R E P O R T R E S U M E S

ED 003 502 56

A STUDY TO DETERMINE THE RELATIVE EFFECTIVENESS OF THE USE OF A SERIES OF FILMED DEMONSTRATIONS IN TEACHER EDUCATION FOR A NEW HIGH SCHOOL MATHEMATICS CURRICULUM.

BY- BEBERMAN, MAX VAN HORN, CHARLES

NATIONAL EDUCATIONAL TV AND RADIO CENTER, NEW YORK
REPORT NUMBER NDEA-VIIA-158 PUB DATE 60
ILLINOIS UNIV., URBANA
GRANT OEG-720010
EDRS PRICE MF-$0.18 HC-$2.68 67P.

DESCRIPTORS- EFFECTIVE TEACHING, *FILMS, *TEACHER EDUCATION, MATHEMATICS CURRICULUM, HIGH SCHOOL STUDENTS, *DEMONSTRATIONS (EDUCATIONAL), *TEACHER ATTITUDES, ACADEMIC PERFORMANCE, *INSTRUCTIONAL FILMS, INSTRUCTIONAL INNOVATION, COMPARATIVE ANALYSIS, URBANA, ILLINOIS

THIS PROJECT WAS DESIGNED TO DETERMINE A SPECIFIC APPLICABILITY OF USING CLASSROOM DEMONSTRATION FILMS AS PART OF TEACHER EDUCATION PROGRAMS, INSTEAD OF THE CONVENTIONAL TECHNIQUE OF LIVE CLASSROOM OBSERVATION. A SERIES OF 20 FILMS WAS PREPARED FROM FOOTAGE OBTAINED DURING THE COURSE OF 77 MEETINGS OF A NORMALLY SCHEDULED CLASS IN THE UNIVERSITY OF ILLINOIS HIGH SCHOOL DURING 1 ACADEMIC YEAR. EACH FILM WAS 30 MINUTES IN LENGTH AND WAS DEVOTED TO ONE TOPIC FROM A COURSE IN BEGINNING ALGEBRA. THE FILMS WERE USED IN TRAINING A GROUP OF 34 TEACHERS AT A SUMMER CONFERENCE HELD AT THE UNIVERSITY. THE ATTITUDES OF THESE TEACHERS AND THE SUBSEQUENT PERFORMANCE OF THEIR PUPILS WERE COMPARED TO THOSE OF ANOTHER GROUP OF TEACHERS TRAINED AT A SIMILAR CONFERENCE HELD THE SAME SUMMER IN WHICH DEMONSTRATION CLASSES WERE TAUGHT BY AN EXPERIENCED TEACHER. THIS REPORT DESCRIBES THE FILMS, THE MANNER IN WHICH THEY WERE MADE AND USED, AND THE OUTCOMES OF THE COMPARISONS MADE OF THE TWO GROUPS OF TEACHERS. SUBJECT-MATTER COMPETENCE OF THE PUPILS TAUGHT BY THE TWO GROUPS WAS MEASURED BY 18 QUIZZES AND 3 UNIT EXAMINATIONS ADMINISTERED THROUGHOUT 1 SCHOOL YEAR. AN UNFORESEEABLE CIRCUMSTANCE WAS THE PLACEMENT OF ALL FOUR OF THE LEAST CAPABLE CLASSES IN THE ENTIRE COMPARISON SAMPLE IN THE HANDS OF TEACHERS TRAINED AT THE FILM OBSERVATION CONFERENCE. WHEN THE DATA FROM THESE CLASSES WERE INCLUDED, ONLY 1 OF 21 COMPARISONS SHOWED SIGNIFICANT DIFFERENCES BETWEEN THE GROUPS. FORECASTS OF DIFFICULTY OF TEACHING SUBJECT UNITS MADE BY THE FILM-TRAINED TEACHERS WERE FOUND TO MORE CLOSELY MATCH ACTUAL EXPERIENCE THAN FORECASTS BASED ON THE LIVE-CLASS DEMONSTRATION. (JC)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

A Study To Determine The Relative Effectiveness Of The Use Of A Series Of Filmed Demonstrations In Teacher Education For A New High School Mathematics Curriculum

Report Prepared by:

Charles Van Horn

Title VII Project Number 158
National Defense Education Act of 1958

Grant Number 720010

Principal Investigator:

Max Beberman,
Director,
University of Illinois Committee
on School Mathematics

Grantee:

The National Educational Television and Radio Center
[with The University of Illinois]

The Research Reported Herein Was
Supported by a Grant From The
U.S. Department of Health, Education, and Welfare
Office of Education

ED 003502

I. Introduction

The University of Illinois Committee on School Mathematics was formed in 1951 by three groups within the University, each concerned in its own way and for its own reasons with the improvement of mathematics education. The College of Education, mindful of its professional responsibility to teachers and students of mathematics, believed that by revising the role of the teacher and providing him with improved materials, the emphasis in secondary mathematics instruction could be shifted away from acquisition of skills and toward mastery of mathematical principles. The Department of Mathematics was concerned with bringing high school courses closer in spirit and mathematical outlook to that of more advanced courses in the field, and the College of Engineering, as one of the principal consumers of the output of high school mathematics classes, saw in the Committee a means of increasing the total amount of mathematics studied in high school so that engineering freshmen could be exposed earlier in college to professional courses.

For several years the Committee devoted itself principally to two interlocking objectives: the preparation of a set of text materials which would set forth, in teachable form, the basic concepts of mathematics, and the establishment of a set of teaching procedures and devices adapted to such material. The pursuit of these objectives has been, in practice, a single venture. The organization of the material and the form in which it is presented to the students have been altered many times over since the publication of the original text in 1952, but each alteration has undergone the careful scrutiny of a professional mathematician whose concern is to insure that mathematical rigor is never sacrificed to pedegogical expediency. On the other hand, no principle, regardless of its mathematical importance, has been incorporated into the texts without regard for the extent of the pupil's knowledge or the state of his conceptual structure of mathematics. No topic is ever introduced in a UICSM course unless the pupils are equipped to understand its underlying principles and to order it to an appropriate place in their own understanding of the structure of mathematics. These practices insure that the growth of the pupils' ideas of what mathematics is about is an orderly one and that mathematics is viewed as a well-regulated and integrated structure rather than as a collection of unrelated problem solving devices.

For the first few years of its existence, the problem of teacher training was not a serious one because the program was taught only in a few schools by a handful of teachers, all of whom were either members of the project staff, or were visited frequently by staff members. As the program found acceptance, the problem arose but was still of minor concern because personal attention could still be given to a small number of new teachers, even though they were widely separated geographically. Continued growth made the problem more urgent; the amount of personal attention that could be devoted to each teacher dwindled until finally, in 1959, in the face of increasing demand for trained teachers, the Committee was forced to include in its professional concern, the need for instituting a teacher training program.

The need for specialized training and the kind of training that is required may best be understood by examining the nature of the materials with which UICSM teachers work. A UICSM course differs from a mathematics course conventionally taught at the same grade level, both substantively and procedurally; the substantive differences are in the inclusion of topics not ordinarilly treated before college level, such as the notation and procedures of set theory, and the axioms on which algebra is based; the procedural differences include the implied attitude toward the careful use of language, the source of the pupils' ideas, and the internal consistency of the subject matter. The experiment described here is concerned entirely with the procedural aspects of UICSM teaching, and not at all with its substance.

Concern for the use of language is demonstrated by the UICSM in the manner in which it is used, both by teachers and by pupils, and by the teacher's attitude toward the pupils' utterances, whether spoken or written, and whether they are couched in verbal or mathematical symbols. In the face of careful phrasing by the text and the teacher, and as a consequence of the latter's responses to his own efforts, the pupil soon finds that he too is expected to exercise the same degree of care, although this admonition is never explictly made. The acquisition of such sensitivity toward language and the appreciation of this quality in others are lessons which the UICSM hopes to teach to each of its pupils.

The second respect in which the atmosphere of a well-taught mathematics class differs under the two programs is in the source from which pupils receive their ideas. In a UICSM text the most important ideas are left unstated when they are initiated so that the pupil will have no recourse but to derive these for himself. This attitude does not imply dissatisfaction with verbal language nor an attempt to impede communication between the teacher and pupil, but reflects the conviction that the pupil understands more thoroughly and retains more completely those things which he has discovered for himself. A new principle is introduced by exposing the pupil to repeated instances of it in a series of exercises, either in his homework or in a classroom development. Such a series is carefully devised so that the problems, which were quite easy at the beginning, become increasingly difficult until the learner is prompted to seek a new method of attack or a short-cut. Provided with this incentive, the pupil detects the pattern inherent in the work he has done to that point, perceives the operation of the new principle and, without having made a formal statement of it, utilizes his knowledge to finish the series. Only after he has had the opportunity to become thoroughly familiar with that principle and has learned its applications, limitations, and implications, is he required to formalize or to verbalize the principle he discovered.

The third difference between the attitude of the UICSM toward learning mathematics and that displayed by teachers and conventional texts is in the nature of the organization of the subject matter and the degree of concern expressed for that organization in the

development of the course. Since most algebra classes are taught without regard for the deductive organization of the subject, it is not possible to interrelate all of the material they present. Various topics are introduced as relatively discrete units, each to be mastered and then laid aside for the next, until the pupil comes to regard algebra as a collection of loosely organized tricks, largely independent of one another, by means of which problems may be solved. If, by contrast, a pupil begins his study of algebra with a set of principles that are described as basic and, from that point on, is required to relate each new principle to the basic ones or to others which he has derived from the basic ones, he is shown that all of algebra exists as an integrated and consistent structure within which arbitrary and peremptory decisions are never permitted.

The first two of these aspects of the UICSM procedures, careful use of language and the utilization of discovery, rely to a considerable extent on the behavior of the teacher for their successful application; the third inheres in the text and is less subject to teacher manipulation. Although the text is prepared with great concern for the precision of its language, a teacher who is careless in his own statements or tolerates carelessness in those of his pupils outweighs its influence. A teacher who over-verbalizes and, in so doing, reveals prematurely to the class those ideas which they should be permitted to discover for themselves, will never be effective in teaching this material.

The nature of the training that is required to prepare a teacher for such a program should be clear from this description of its nature. An obvious need is mastery of the subject-matter by the teacher. The UICSM First Course (beginning algebra) contains much material that has traditionally been treated only in advanced college courses in mathematics and is unfamiliar to many high school teachers. This material is not difficult, but it represents ideas which they have never encountered, ways of thinking they have never experienced, and subject matter which is unfamiliar to most of them. The greater portion of the time spent in the summer training sessions conducted by the UICSM is, therefore, devoted to mastery of the subject matter to be taught with special concern for their understanding of its deductive organization. The teachers must master in four weeks the same material with which their pupils will spend an entire school year; this requirement, although it is a demanding one, has never presented a major obstacle in training.

The second requirement in training teachers to use the UICSM material is difficult to describe -- and even more difficult to accomplish. Many teachers entertain doubts concerning the ability of their pupils to master such subject matter at this grade level and, for these teachers, proof is necessary of the ability of ninth graders to perform satisfactorily with the First Course. To other teachers, steeped in the traditional concepts of algebra teaching, the principles of discovery and delayed verbalization seem strange and self-defeating; for all teachers, the classroom strategies by which these principles are utilized need to be explicated and demonstrated.

All of these requirements are satisfactorily met by the observation of a class in session under the direction of an experienced teacher. Such observation provides first-hand evidence of the ability of the class to make discoveries, to utilize them without having made an explicit statement, to generalize from their own experience, and to profit by the emphasis on structure and pattern that the UICSM substitutes for drill and memorization. It provides the teachers with an opportunity to observe, in context, the kind of behavior by which the teacher directs the class' thinking, adduces generalizations without stating them, overcomes semantic obstacles without overtly recognizing them, and judiciously exploits the curiosity and intellectual agressiveness native to children at this age.

Classroom observation has been included as a part of all UICSM teacher training conferences because no adequate substitute has ever been recognized for it, but its advantages are not available without a price. For a school administrator, the process of convening a demonstration class during a summer session is not only difficult in itself, but the head start that such a class affords to a small number of pupils carries the difficulty forward in time throughout the ensuing school year. From the standpoint of the pupils involved, the conditions under which such a class recites are not favorable; the presence of an audience seated behind them inhibits the behavior of some pupils and prompts others to display more concern for the reactions of the audience than for those of the teacher.

From the standpoint of the observer, the usefulness of the observation sessions is limited in two ways. Their view is restricted to the teacher and the back of the pupils' heads, depriving them of the point of view from which they will be teaching and of the information that an alert teacher finds in gestures, facial expressions, etc. Their observation is further limited to a few sessions (usually twenty or less) which, in turn, limits both the range of classroom situations that might occur and the extent to which the longitudinal development of topics is possible. It is not unheard of for a teacher whose training is based on the observation of such a demonstration class to report that for the first twenty days (the length of his observation) his own class proceeded normally enough, but that on the twenty-first day, when he no longer had detailed notes to which to refer, some trouble was encountered.

Faced with a constantly increasing demand for enlargement of its teacher training program and aware of the limitations of its existing procedures, the UICSM in 1958 agreed to work with the National Educational Television and Radio Center in preparing a set of motion pictures which would portray a First Course class in session over an extended period of time.

The existence of such a series of films would make it possible to multiply, by any desired factor, the availability of teacher training facilities without incurring the administrative and social liabilities inherent in live class observations. An even greater advantage that might accrue from the use of such films for this purpose is the possibility that they offer of condensing, into a few screen minutes, the events that occurred over a period of many

weeks or months of classroom time. It was hoped that the basis on which the trainees' observations were founded could be broadened and, by judicious editing, that the emphasis on longitudinal development of the subject matter could be sharpened and made more apparent to those teachers to whom the films are addressed.

The production of the films and their applicability to this kind of teaching raised questions of such a nature that both of the groups involved hesitated to attempt so fundamental a change in the established pattern of teacher training without evidence that nothing of value to the trainee was sacrificed. Administrative convenience was not deemed sufficient compensation for possible losses in instructional efficiency, and postulated gains through time-compression and longitudinal development had never been demonstrated. For these reasons, a comparative study was undertaken by which the relative effectiveness of observations of a live class and of the films could be compared for their usefulness to the teacher-trainee and for the 'second-generation' effects they might have on the performance of the pupils of teachers trained by these methods.

The study was made possible by a grant from the U.S. Office of Education under the terms of Title VII of the N.D.E.A. (P.L. 85-864). Grant Number 720010.09 was made to the National Educational Television and Radio Center which, in turn, sponsored the production of the films and made funds available with which to conduct the comparative study reported here. The experiment was conducted with two teacher training conferences held by the UICSM at the University of Illinois during the summer of 1960 with the financial assistance of the National Science Foundation.

Each film is a half-hour in length and deals with a single topic. The films do not attempt to teach the subject matter with which they deal; in fact, familiarity with the subject matter is a necessary condition and they are arranged in such a way that, with one exception, the teachers will have studied the subject matter of each film before seeing it.

The twenty titles of the series and a precis of each film are shown below.

I. Introduction to Real Numbers

The class becomes aware of the need for numbers which designate direction as well as magnitude, and invents numerals to designate such numbers. This film begins with the third lesson of the course and is distributed over four days of work.

II. Adding Real Numbers

The class learns to combine real numbers and formulates rules for combining them. In a latter sequence, they are shown two months later as they compose and state rules for the ideas they have known and used for ten weeks.

III. Advent of Awareness

This is one of two films of the series which does not include classroom footage. Several examples of primitive learning are shown; one human, three canine, three equine, and several pertaining to an experiment involving a white rat.

IV. Multiplying Real Numbers

The class is shown as they encounter, in their seventh lesson, a situation in which combinations of three real numbers are involved. With the help of a conceptual model, they find short-cuts for these problems and, in the forty-ninth lesson they are shown composing and stating the rules for such operations.

V. Operations and Inverses of Operations

The class encounters the concept of a singulary operation (a set of ordered pairs) in lesson 26, finds in lesson 38 that all singulary operations have inverses, and formulates a definition of 'inverse' in lesson 73.

VI. Subtracting Real Numbers

The first subtraction questions are encountered in lesson 4, the concept is made explicit five weeks later and, after five more weeks, the rules are stated precisely.

VII. Dividing Real Numbers and Ways of Naming a Quotient

As early as lesson 7 the way is being prepared for learning the operation, division. It is dealt with in detail in lesson 35. In lesson 76 the teacher departs from the text to exploit a lead opened by a student's remark and a logical basis for division is established.

VIII. Principles of Arithmetic for Numbers of Arithmetic

Several brief sequences covering a period of two weeks are used to show how the class becomes acquainted with the ring properties of numbers of arithmetic and verifies discoveries they have made concerning commutative and associative principles.

IX. Distributive Principles for Numbers of Arithmetic

In lesson 13 examples are shown which promote awareness of the principle; in lesson 18 it is practiced, named, and its pattern constructed.

X. Principles of Arithmetic for Real Numbers

In lesson 24 the class attempts to apply the principles of arithmetic to operations with real numbers; the additional principles that are necessary emerge in lessons 29 and 30; in lesson 48 the principles are stated and, in lesson 50, the postulate system is completed.

XI. Comparing Real Numbers: The Number Line, Distance, Absolute Value

In lesson 36 the class confronts the relation, less than, and devises a number line to test it. As a consequence of an attempt to determine the meaning of 'distance between real numbers', the concept of absolute value is discovered and named.

XII. Prerequisite to Communication

The second film in the series which does not include classroom footage, utilizes commentary and examples of nonverbal learning in an examination of the nature of communication and its pedagogical implications.

XIII. Prelude to Deduction

Shows some of the preparation that is made to increase the pupils' confidence in formal proof. This preparation, together with the principles of arithmetic (Film X), leaves them ready to undertake proving generalizations.

XIV. Proving Generalizations

The class is shown as they learn how to construct test-patterns and to use them to verify generalizations. A sample of their performance is included.

XV. Principles and Discovery in Algebraic Manipulation - I.

After the principles are mastered, the class is seen using the distributive principle in lessons 59 and 60 to do what would conventionally called 'removing factors' and 'collecting terms'. This is their first experience with manipulation.

XVI. Principles and Discovery in Algegraic Manipulation - II

Further instances of the kind and use of manipulative work begun in Film XV.

XVII. Sentences, Roots, Solution Sets

Another class is seen working with graphs of sentences under a teacher who has not taught this material before. In a later sequence, the regular class is shown working with a teacher they have never seen before.

XVIII. Equivalent Equations and The Transformation Principle

The entire film was made from footage obtained in a single class session (Lesson 100), but it rests on many bases. The class had just acquired the concept of subset of a set and had been working with uniqueness and cancellation principles and with sentences having the same solution set when they were photographed as they encountered an equation they could not solve by inspection.

XIX. Using Transformation Principles

The class is shown in their first encounter with an equation containing fractions in lesson 100 and again when they encountered certain complications the next day. The final sequence taken from lesson 130, climaxes the work on inequations.

XX. Solving Worded Problems

The class is shown first, just after they have learned to use pronumerals and again, five days later in the process of finding answers by inspection. Nine weeks later they approach the task of converting words into equations in a more systematic manner.

II. Production of The Films

At the time this comparative study was planned there was no assurance that it could be completed because of the formidable technical problems that confronted the members of the staff responsible for turning out the films on which it was to be based. No precedent was available for the production of films under these circumstances, and the possibility was recognized from the outset that a series which conformed to the requirements set up by the project staff could not be achieved.

Classroom made films are ordinarily the product of a compromise; either a classroom is constructed and the class conducted in a physical setting which conforms to the requirements of the production staff, or an existing classroom is converted into a makeshift sound stage. Either of these procedures is suitable when a small amount of film is required, but neither can be tolerated when, as in a case of this kind, plans are made to photograph all or part of as many as eighty class sessions in a school year. In either case, the presence of cameras, camera crews, lights, conspicuous microphones, etc. cannot help but affect the behavior of the class and will almost certainly detract from the level of performance of the pupils. It was exactly this effect which was uppermost in the minds of the planners of this film series; the fundamental requirement laid down in preparation for these films was that the presence of cameras and their related equipment must never be allowed to intrude upon the activities of the class or the teacher. Meeting this requirement was expected to be a considerable obstacle to the functioning of the film crew but, as a result of careful planning, good rapport between the film crew and the mathematics staff, and the cooperation of the pupils, few if any difficulties were encountered.

The first step in planning for these films was a general survey of the technical problems expected to be encountered and a discussion of these with members of the production staff of the Department of Cinema at the University of Southern California. Their suggestions with respect to the type of equipment to be used and their advice concerning the procedures by which many potential problems might be avoided, were important contributors to the success of the series.

The members of the production staff were obtained from the same department on the basis of recommendations by the faculty. This staff was headed by Byrl L. Sims, who was assisted by Stanley Follis and John Werner. This group supervised the installation of the necessary facilities, made revisions and adjustments in the general plans, and devised the procedures by which the films were edited.

Equipment and Procedures

The physical setting in which the films were made was provided by making alterations in an existing classroom in the University of Illinois High School. Camera booths were constructed, acoustical tile was installed on the ceiling, rear wall, and part of the front and side walls of the room, and fluorescent lights replaced the existing room fixtures.

The largest camera booth extended across the width of the room at the front and provided space for a camera, a part of the equipment necessary for sound recording, operators for both of these and, if necessary, one observer. The camera in this booth operated through a very large port located just above the chalk board so that its coverage included the entire classroom and any pupil could be photographed from this position. This freedom became important to the tactics of directing and editing since it made possible the detailed close-up coverage of individual pupils which contributes to the spontaniety and feeling of presence that characterize the films.

Another booth across the rear of the room existed primarily to house the camera which provided continuous coverage of the teacher and a record of the chalk-board. It served the additional function of providing space for as many as three observers (for whom an additional glass port was provided) and, by virtue of the angle which it makes with the front wall, provided an acoustical advantage.

A third booth at the side of the room housed a camera which could cover any of the class except those located at the rear of that side of the room, as well as the corner and side not accessible to the rear camera. This position made it possible to provide continuity of film coverage of the teacher when he took up a position in that corner or approached the desk of a pupil located on that side. It also provided the editor with cut-away sequences by which it was possible to avoid the confusion or monotony that might result when scenes made from two locations almost exactly opposite one another are alternated. This side booth was just large enough to house the director, the programmer, and the recorder on which the sound track was recorded. Each booth was provided with a monitor speaker to enable the occupants to hear that which took place in the class room.

All of the cameras were 16 mm. Mitchells, equipped with synchronous motors especially wound to permit continuous operation for periods as long as one hour, and mounted on Mitchell Hi-Hat mounts which were secured to the walls of the camera booths. Each camera was provided with a magazine which could accomodate a two thousand foot spool of film, and with a Pan Cinor Zoomar lens with focal length continuously variable from 17.5 to 70 mm; this permitted a smooth transition from wide-angle coverage of the entire class to selection of a single pupil for a close-up picture. These lenses require refocusing when an alteration is made in their effective focal length, a situation which presented the cameraman with a task beyond his physical capability in view of the size and shape of the camera. The problem was resolved by assigning the task to an assistant stationed at each camera. The correct lens positions for a number of previously agreed upon locations in the room was inked carefully on the lens barrel and the assistant focused to these in accordance with oral instructions from the cameraman. The only other problem encountered in camera operation, the occasional failure of the take-up mechanism to function in the early minutes of a session when a heavy spool of film was in the feed roll, was solved by fitting a larger take up spool to provide greater torque.

Lighting for all interior shooting except inserts was provided by ceiling mounted fluorescent tubes which, in conjunction with the row of ceiling-high windows along the east wall and the high reflectance of the white walls and ceiling, provided about 110 foot-candles of illumination at the level of the pupils' faces. The class met in the early afternoon so uniform lighting, without glare or shadow was readily obtained. The only supplementary lighting used was that made necessary by a shadow cast on two pupils in the front row by the camera booth which covered part of two windows. The boost provided by two fluorescent desk lamps (a total of sixty watts) placed on top of that booth and directed toward the shadow area was sufficient to bring its level to that of the rest of the room.

The entire series was shot using Eastman Tri-X Reversal film with a shutter opening of 235°, corresponding to an exposure of 1/36 (to be precise) second, and with lens apertures which varied from f. 4. 5 to f. 5. 6 according to camera location and local atmospheric conditions.

The most difficult technical problem encountered in the entire production was that of sound recording. The original plan called for the use of three omni-directional microphones (Telefunken, condenser type) suspended at ceiling height, approximately at the vertices of an equilateral triangle, for picking up pupils' recitations and a wireless breast-pocket microphone (Stephens Tru-Sonic) for the teacher. All of these fed through a mixer (Altec 230-B) and the mixed signal was led to the magnetic film recorder (Magna-sync, 916-10). The ceiling, the entire back wall, and a part of the front and side walls were covered with acoustical tile and a carpet was placed on the floor to minimize sound reflections. Tests showed that this arrangement did not yield adequate sound tracks, largely because of the wide range of voice levels among the pupils and the exceptionally heavy voice of the teacher. The quality of the sound was brought to an acceptable level by eliminating the teacher's microphone and one of the ceiling mounted ones, and by occasional reminders to the class concerning the effects of mumbling, pencil tapping, opening note book rings, etc. The most useful kind of admonition was found to be the playback to the class of a few feet of one of their previous performances, especially if it were characterized by indistinguishable voices and identifiable notebooks.

Later alterations brought about two further noticeable improvements in sound quality; the room was further deadened by covering the ceiling and most of the remaining wall space with glass wool batting, and the faint hum, once thought to have been the result of some characteristic of the electronic equipment, was eliminated by redesigning the camera mounts. These mounts were originally secured to the front walls of the camera booths and the faint vibrations which they transmitted to that wall was audible in the recording during moments of room silence. When the cameras were moved to mounts bolted to the concrete floor, no further hum was recorded.

No serious attempt was ever made to isolate the room from external sound sources such as traffic in the street below or the noises from steam lines; these are occasionally heard in the films, but do not consitute an interruption or distraction.

Editing

The process by which the films were edited into their final form was, by necessity, devised by the production staff to fit the needs of this series. The problems raised by the nature of the films, the amount of film footage handled, and the rate of production were such that conventional procedures would never have succeeded.

The first circumstance that dictated a major revision in editing procedures was the lack of a formal script for any of the films. They were, quite literally, written at the editing table after all of the footage was processed. Their content, in terms of intended subject matter, was determined by the staff of the mathematics project before the school year opened, and the advisers to the film unit were able to predict, with reasonable accuracy, the date on which a desired point would arise in a class session. The teacher was able to exert some control over the time of occurrence of a desired point and much greater control over the form which the class discussion would take. Within these limits, the appropriate material was to be found buried in the day's footage and it became the task of the editors to identify the sections specified by the programmers as relevant to the theme of one of the films, and to fashion from this raw material, a film which would fulfill both the technical demands of the mathematics staff and the formal requirements of the film medium.

The second circumstance which necessitated the revision of editing procedures was the unprecedented volume of film handled. The simultaneous operation of three cameras consumes 108 feet of film per minute, so if a fifty minute class session is photographed in this way, 5400 feet of film must be processed. In making the eighteen classroom films of this series, 77 class sessions were filmed in whole or in part, totalling 2950 minutes of camera operation, and a total of approximately 400,000 feet of film, and more than 100,000 feet of sound were accumulated. The use of procedures designed for production situations in which the ratio of exposed film to final length varies from 4:1 to 10:1 would have increased the time needed to finish this series by a prohibitive extent.

In general, the editing procedure used to meet the joint demands of this tremendous volume of film and the loose degree of organization of script planning was based on a series of successive approximations to final form. When a day's film was returned from the laboratory, it was screened from three projectors simultaneously and with synchronized sound. Two purposes were served by this session: the production crew was able to evaluate the technical quality of their work so that deviations from acceptable quality could be corrected before they became serious, and the mathematical staff was enabled to note the sections that were considered suitable for inclusion in one of the films.

The footage selected in this way then went through a preliminary editing phase in which the editor removed that part of the original not designated for use in one of the films. On the remaining original, the two rolls containing coverage of the students were edited into one roll;

work print was then made on this roll along with the corresponding roll which showed the teacher. To have made work print from all of the footage would have increased both the expense and the volume of film beyond any possibility of efficient operation.

With the preliminary decisions made and the work print ready, the editor was able to undertake the task of converting the three sets of pictures into a single roll. His specific chore at this point was to combine these, according to the notes supplied by the programmers, in such a way as to utilize most effectively the visual potential of the three scenes. With the exception of those points at which the material being written by the teacher at the board was of primary importance, the editor was given complete freedom to carry out this task according to his own judgment and his knowledge of the techniques and effects of editing. The specific visual content of each film, choice of scenes, selection of points of view, and the location of transition points from one to another were entirely in the hands of the editor at this stage.

It was in connection with this consolidation of rolls that the greatest demands were placed on the adaptability and ingenuity of the production staff, for it was at this point that the effects of the large quantities of film handled were felt most keenly, and the decisions made during this process had a considerable effect on the acceptability of the films to the audiences. The process was expidited by a device which incorporated three Zeiss Moviescope viewers and a sound reader into a single editor's viewer. This arrangement permitted simultaneous viewing of all three rolls and provided the editor with a maximum of information and the freedom with which to utilize it.

The initial screening of the first rough-cut version of a film was an important one in its conceptual development. The three screen viewing of the daily footage was useful and important, but it suffered from several shortcomings; it placed unreasonable demands on the perceptual ability and attention span of the audience, and the programmers in particular were further restricted by the necessity to evaluate the material, in terms of its subject-matter suitability to one or more films, as they watched it. To these programmers, who were somewhat less than sophisticated in their cinematic expectations, it provided the first definitive idea of the nature of the finished film and transferred their task into a new domain of decisions.

The first rough-cut version of each film was characteristically from two to four times the length of the finished film; the process by which it was brought to its desired length was the elimination, by successive viewings and careful review of its content, of those sections which could most readily be spared. The judgment of the editor was required in this process in order to maintain the unity and integrity of the film with respect to its visual properties, but less freedom was available to him in this stage because, with each deletion, the nature of the remaining material was altered and the emphasis placed on the careful selection of content was increased.

After the film had been reduced to the appropriate length, explanatory comments were prepared to introduce separate sequences, orient the viewers to the subject matter to be covered, point out the passage of time, since some of the films cover time spans as long as three months, or to summarize the essential elements of the material demonstrated in each film. The comments were prepared by the mathematics staff of the film project and were narrated by the teacher, who was photographed in a separate setting to avoid confusion between his two roles.

The steps involved in shooting these inserts, cutting the negatives, synchronizing sound, etc. were routinely carried out and need not be elaborated on here.

Programming

In a project of this kind the selection and organization of the material to be presented are obviously of primary importance. To compress a full year of classroom work into eighteen half-hours of screen time requires that a full ninety-five percent of the class time be sacrificed and, inasmuch as approximately a fourth of the total screen time is devoted to other-than-classroom footage, the ratio of depicted to undepicted classroom time becomes about 1:25.

One of the greater advantages postulated for this series was the possibility it offers of tracing the development of an idea longitudinally as it appears in its various forms and contexts; this advantage can be realized only if the appropriate class sessions are identified in advance and recorded on film. The selection of the class sessions to be photographed and the organization of these sequences into films was the responsibility of the programming section. Two mathematics teachers were assigned to the film project for this purpose. Their tasks, the specific nature of the problems that confronted them, and the procedures by which these problems were resolved are matters too complex and technical to be detailed in this report, but an outline of these activities may clarify the nature of the films.

The activities of the programmers varied with the status of the production task; in the planning phases, programming consisted of the specification of topics to be treated; during production it required identification (prediction) of the class sessions in which the specified topics would be discussed and the selection of the obtained footage with which to illustrate them; during the utilization of the films, the same programmers were assigned to the preparation of a descriptive manual to be distributed as a supplement to the films.

The planning and organization function of the programming task may be thought of as consisting of five steps which, although they are separately described, were by no means discrete. The first step in planning the series was the selection of a set of topics and ideas which are important to the First Course and which may be expected to be encountered by any teacher. These were grouped under a smaller number of descriptive headings which were, in a primitive form, the specification of the content of each film of the series and which, in many instances,

became the title of the film to which it pertained. The selection and grouping were done by subject-matter experts and experienced teachers without concern for visual content, time requirements, or any of the operational details involved in the film-making process. After this task was completed, the ideas were translated into terms of page numbers from the text to enable an estimate of the amount of material to be included under each heading. These page number outlines constitute the closest approach that was made to a shooting script, and it was from these that the programmers and film crew worked during the initial steps of the implimentation phase.

The second planning stage was begun after the decision to undertake the experiment had been made and the physical preparations started. At that time, an experienced teacher, Miss Gertrude Hendrix, was assigned to the post of programmer for the film group; throughout the duration of the study she was assisted in this task by another teacher from the project staff, Mr. Jesse A. Orvedahl.

This phase of the work consisted of the preparation of an extensive and detailed set of notes based on the programming teachers' experience and their observations of a class then in progress. These notes dealt with the material in the basic outline, but went far beyond it in depth and scope and, although they made specific reference to the films, they were more concerned with teacher and pupil behavior than with visual matters. These notes were to assume considerable importance to the teacher of the film class in shaping individual decisions that were required as the year progressed.

The third major step in the programmers' preparations was a precautionary one. A class studying the same material as the film class under the same teacher was scheduled and treated in such a way that they remained about two weeks ahead of the class to be filmed. By observing the meetings of this class, the programmers could refresh their memories concerning the basic outline, supplement their notes, and try out procedures and teaching techniques that were tentatively selected. The two week interval was selected in recognition of the film unit's requirement that specification be made at least a week in advance of a class session to be filmed; in this way, the programmers were able to devote nearly a week to the assimilation of the information gained from observation of this class.

The phase of the programming, described here as the fourth step in the series, began with the preparation of the extended notes and continued throughout the preparation of the printed supplements. This work, the preparation of a large and relatively elaborate card index became the organizational nucleus around which revolved the organization of the subject matter, the prediction of pertinent class sessions, and the identification of useful film footage.

Seven card files, four basic and three derived, were prepared and maintained under the titles: Concepts, Generalizations, Events, Films, Lessons, Interpretive Comment, and Manual.

In the first two, Concepts and Generalizations, a card was alloted to each of the ideas (appropriately classified) that appeared in the basic outline and the programmers' notes. Each file was identified by card color, and included a reference to the function of that idea in the course, according to its intent to develop an awareness of that concept or generalization, or to make an effort to name or state it. Greater emphasis was placed in preparing this file on comprehensive coverage of the course content than on the probability of film coverage or the corresponding visual content, since it was considered more profitable to make a restricted selection from full coverage than to attempt to pad a limited repetory.

The Interpretive Comment file, again identified by a characteristic card color, was built from questions, comments, or suggestions made in staff conferences, by consultants, by audiences in pre-release showings or in classes that were observed or recalled. This file, like the others, was built without concern for film content, since the material it contained was equally applicable both to the films and to the printed matter which accompanies them.

The Events file, contained cards which bore quotations, comments, and narrative descriptions of incidents which have occurred in First Course classes and which, if they were to reoccur and be recorded on film would be useful in developing the theme of one or more of the films. The cards in this file were organized according to the section, in twenty page units, of the First Course text.

These four files, along with the basic outline of content constituted the source materials from which the ideas contained in the series would be drawn and organized. The process by which this was to be accomplished was the function of the remaining three.

A file labeled Films contained a duplicate of each card from the first four, organized according to the title of the film to which it pertained. This precaution was made necessary by the intent of the series to extend and compress time in the editing of the films; a single film might be made from footage obtained during one class session, or it might contain footage from as many as ten classes separated in time by as much as three months, and reference to this file clarified the problem thus created in designating and allocating content. In selecting from showings of daily footage, this file not only placed at the programmers' fingertips the advantage of the distilled thought and experience of the entire staff, but also provided a running record of the status of each film during its production. Before any footage was obtained, the file for each film contained all of the material that that film would include if everything could have been recorded. As editing progressed, the cards representing each point included in the film were removed to an inactive file so that, when a final version was attained, the remaining points could be allocated either to the narrator's comments in the film or to the published material. Further utilization of this film was made possible by cross-indexing it to the chronology of the class' progress so that it became possible to predict, for each point to be covered, the approximate date on which the footage for it would be available.

The fifth file, Lessons, also consisted of duplicates of cards from the basic four but they were organized according to the class sessions in which that concept, event, etc., was expected to appear. When unexpected circumstances brought about a deviation from the schedule, the corresponding cards were simply shifted in the appropriate direction to bring the record into correspondence with the occurrence. This file became both an annotated history of the class' progress through the course and a source of realistic predictions concerning the shooting schedule. The principal value of this file, apart from its continuing usefulness to the authors of the text and its contribution to the expectations of inexperienced teachers, was the maintenance of the index dates in the Films file.

In use, these files constituted both an aid to memory and to organization. Short range predictions of shooting dates, up to about two weeks, were reasonably accurate and day-to-day coordination of class and film crew was closely controlled. Constant reference to the Films file kept the programmers immediately aware of the class topics for which film coverage was considered important; when the Lessons file indicated that one of these was about to arise, the film crew was alerted and the teacher of the class was extensively briefed on the nature of the hoped-for activity and its probable place in the lesson.

On these occasions, the programmers would take the cards pertaining to that lesson from the four basic files, Concepts, Generalizations, Events, and Comments; from these, plans were made for the day's shooting including the sections of the lesson to be filmed, their probable use in the series, their relative desirability, etc. Following this conference, the cards would be taken to the teacher of the class and the nature of the lesson and its implications for the film series discussed with him. Armed with this knowledge, he was usually able to plan the day's class session in such a way that the desired events were achieved. After this conference, and just before the class began, the programmer and the director of the film crew would discuss the plans together so that adequate film coverage of the pertinent sections of the class could be obtained.

The actual class meeting and filming were almost anti-climactic after these preparations. The teacher was wholly in charge of the recitation and the conduct of the class. Communication between him and the remainder of the group was possible only by gesture or by an outright interruption of the class. He depended on his own knowledge of the subject and his skill at controlling the direction and rate of the class' progress to arrange for detailed coverage of those sections of the recitation which were needed for the film series.

The steps taken after the shooting have already been described.

The Class

The class which was photographed in the preparation of this set of films was composed of fifteen regularly enrolled pupils in the University of Illinois High School, and was one of four such sections scheduled. The school is operated as a function of the College of Education, rather than by a school district, to provide facilities for the development of teaching procedures and curriculum materials rather than for training teachers. The general level of ability of the pupils in this school is high because the admission policy is a highly selective one based on competitive examinations; the pace is rapid and pupils who enter this school after completing the sixth grade, complete all of the requirements for high school graduation in five years. Because of these circumstances, the pupils seen in these films are not to be considered as typical seventh graders, or as representative of most student bodies.

Their age and previous academic experience would qualify them as seventh graders in most shcools, but their performance on the tests by which they were selected for admission to University High School places them at or beyond the intellectual level of most ninth graders. The fifteen members of this class range in age from not quite eleven to 12½ years; they show a distribution of arithmetic grade placements (based on the California Arithmetic Test) from 8.8 to 11.0; of reading score centile positions (Coop Reading Test) from 42 to 95; and of verbal IQ's (Thorndike-Lorge) from 122 to 150.

Membership in this section of the First Course was voluntary and the pupils knew beforehand that they would serve as subjects in the preparation of these films; permission for the pupils to be photographed was secured from their parents before school began. At no time before or during the school year, was any pupil instructed by any member of the film crew or the High School staff concerning his behavior during the film sessions, other than an occasional request for clarity of speech or stronger voice level. On one occasion, when the recorded version of a statement made by one of the pupils was found to be difficult to understand because of interference from outside noise, that pupil was asked to repeat that same statement in quieter surroundings at another time. This is the only instance of coaching or prearranged recitation in the series. The pupils were not paid nor given any other form of remuneration for their participation in the series.

The pupils in the class were:

Laura Alpert	Douglas Hittle	Christine Robinson
Peter Brown	Charles Kleiner	Adrian Wadsworth
Mary Cavins	Terry Koch	Jane Wall
John Dixon	Helen Manner	John Wetmore
Marcia Greenwold	Susan Newmark	Michael Wiese

Experimental Use of The Films

The first use of the films, that in which the basic question of the comparative study was to be answered, came in the course of two training conferences conducted under the auspicies of the UICSM on the Urbana campus of the University of Illinois during the summer of 1960. Seventy-six teachers were trained during these two sessions; the information obtained from them during their training and from the pupils in their classes during the 1960-61 academic year constitutes the basic data of the comparative study.

The first step in planning for this conference (other than provision for production of the films) was an application to the National Science Foundation for financial support for the summer conference. This step was a break with precedent, for all previous summer training conferences had been conducted without outside aid, and the expenses incident to attendance were borne by the individual teachers or by the schools which employed them. It was deemed advisable to offer financial assistance because it has become practice for colleges and universities offering summer training to offer such aid in amounts that are not feasible without outside support. It was feared that, in the absence of such inducements, this conference would be placed in an unfavorable position in competing for qualified participants; the importance of the decision to be based on the performance of the teachers trained at these conferences was such that every possible precaution, beginning with this one, was taken to insure that these teacher-trainees were well qualified.

When this source of support was assured, the routine announcement of the conference made by the NSF was supplemented by invitations mailed to the administrative officers of the schools participating in the program and to several others from whom inquiries had been received, requesting applications from qualified teachers. This practice was in accord with the UICSM policy of placing the selection of teachers in the hands of school officials who have more and better opportunities than the project staff to observe those teachers and assess their suitability for specialized training. It was specifically required of all applicants for this conference that they submit, with their formal application, a letter from an official of the school system in which they were to teach, indicating that applicants trained at this conference would be assigned at least one class of college-capable eighth or ninth graders to whom they would teach the UICSM First Course, and that the school would cooperate further to the extent of permitting the tests and questionnaires associated with this study to be administered to that class. Each applicant was also asked to specify the dates that would be most convenient for his attendance, and the promise was offered that, insofar as was possible, this preference would be honored. This statement of preferred dates of attendance was made without the applicant's knowledge of the intended nature (control or experimental) of either conference.

Applications were received from more than one hundred qualified applicants, of whom seventy-five were invited to attend the conference; one teacher attended the conference and completed the course of training without receiving a stipend.

In all cases but one, the applicant's preference for date of attendance was followed, a circumstance made possible by the existence of many who had indicated that either session was suitable. A few applicants declined the extended invitation when the invitation made to them conflicted with other plans or committments. The names of the participants in these conferences, and of the schools in which they taught, are shown in Appendix A, together with the identifying numbers, assigned by a random procedure to their class (or classes) for use in processing test data.

In making assignments to the two conferences, an attempt was made to insure reasonable comparability of the two groups of teachers with respect to such characteristics as age, sex, experience, amount of formal preparation in mathematics, and the size and nature of the schools in which they were to teach. Tables 1 to 5 show the comparisons between groups in these respects.

Table 1
Comparison of Conference Groups
With Respect to Age of Participants

	C_1	C_2	Total
N	34	34	68
Mean	37.97	36.24	37.07
S. D.	11.55	10.01	10.82
t	0.65		
R	0.52		

Table 2
Comparison of Conference Groups
With Respect to Mathematical Preparation of Participants
(Semester Units)

	C_1	C_2	Total
N	32	34	66
Mean	36.09	31.47	33.71
S. D.	17.42	14.82	16.30
t	1.14		
R	0.25		

Table 3
Comparison of Conference Groups
With Respect to Experience of Participants
(Years)

	C_1	C_2	Total
N	33	34	67
Mean	11.91	8.09	9.79
S. D.	9.17	8.85	9.17
t	1.73		
p	0.08		

Table 4
Comparison of Conference Groups
With Respect to School Enrollment

	C_1	C_2	Total
N	32	32	64
Mean	895	987	941
S. D.	670	668	670
t	0.77		
p	0.44		

Table 5
Comparison of Conference Groups
With Respect to Sex of Participants

	C_1	C_2	Total
Male	20	22	42
Female	*14	*12	26
Total	34	34	68

*Each conference group included seven Sisters from Catholic schools.

The comparison of sex distributions shown here is that of the 68 teachers from whose classes data were obtained; there is associated with this array, a value of Chi Square which indicates a probability greater than 0.5 of chance deviation from identity between conferences. The distributions of the values on which these tables are based are shown in Appendix B.

In examining these tables, the first step taken was a comparison of the variances of the two distributions for each characteristic. None of these ratios is large enough to warrant rejection of the hypothesis of no difference between groups, even at the .10 level of confidence; from this it is inferred that, in every case, the shape of the distributions is the same for both conference groups and that a comparison of means by t ratio is justified. The t ratios associated with each of these differences and the probability of chance occurrence are shown in the appropriate tables. The number of cases included varies from table to table because not all of this information was available from every participant in the conference.

Examination of these tables reveals that the means of the control group are slightly higher in the comparisons of age, number of units of mathematics studied, and enrollment of the schools in which they teach; these differences are slight, however, and differences as large as these or larger could easily have arisen by chance. The difference between the groups with respect to experience merits more attention. This comparison indicates that the control group possesses more experience than does the experimental group and that this difference is so large that it is difficult to attribute it to the operation of chance factors alone; the t ratio accompanying Table 2 indicates that this difference could be expected to occur not more than eight times in 100 randomly selected pairs of groups. This level of confidence, although it falls short of the five percent expectation commonly required for rejection of a null hypothesis, is large enough to be considered in assessing the importance of its effects on the outcome of the experiment.

The difference between these means is that between approximately eight and approximately twelve years of experience; it could be argued, with some merit, that the benefit gained from added experience depends as much on the total experience as on the amount added, and that the gains in teacher competence accompanying the four years between the eighth and twelfth years of teaching are not apt to be large. This is equivalent to saying that, during his first eight years in the classroom, a teacher gains far more than he may expect to during the next four. This line of argument overlooks the fact that the differences compared are means, and that the experience which is of concern to the pupil is that of an individual; reference to the distributions of experiences, tabulated in Appendix B, casts interesting light on this comparison. The two groups seem to be about equally distributed over the middle two-thirds of the range; the experimental group is clearly over-represented in the category 0 - 2 years, and under-represented in the range 23 - 28 years. Examined in this light, the control group is seen to contain considerably fewer teachers at the level of experience where it becomes an important consideration. Some doubt is cast on the efficacy of the procedure by which teachers were assigned to conferences, and a bias of undetermined magnitude is introduced into the comparison in favor of the control group.

It was not, of course, possible to effect any separation of the groups on the basis of the intellectual ability or mathematical aptitude of the pupils that these teachers were to have in their classes, since no method is known by which a valid prediction of success in First Course can be made.

It was necessary to leave the group imbalanced with respect to geographical distribution because of systematic regional differences in school calenders; it is usual practice in most western states to schedule the school year in such a way that classes end in late May or early June while, in many eastern schools, classes remain in session until late in June. Many of these applicants were from schools along the Atlantic coast which were still in session when the first (control) conference began on June 20, 1960 so these regions are over-represented in the experimental (trained by film observation) conference which opened late in July. This circumstance, which could not have been avoided, is assumed to be a contributor to the difference between the groups with respect to their experiential and academic backgrounds. Eastern schools which use the UICSM program (and these are not necessarily representative of all schools in that region) tend to have slightly larger enrollments, and to contain mathematics teachers who are, on the whole, younger, have less experience, and slightly less formal preparation in mathematics than do their western counterparts.

In both conferences, when the teachers arrived in Urbana, they were immediately oriented to the experimental nature of the summer's work and to the nature and extent of their expected participation in it. They were also told of the nature of their training -- a morning session to be devoted to observational training in classroom method (either live class or a film), a critique of that session, and the remainder of the day to be devoted to a study of the subject-matter of the course.

The subject-matter classes were conducted, in both conferences, by two teachers from the UICSM staff, Miss Eleanor McCoy and Mrs. Alice Hart. This arrangement was made necessary by the size of the two groups; approximately forty teachers attended each conference, so it was possible to constitute two groups of convenient size. In making assignments to classes, the same procedure was followed in both conferences; from an alphabetically arranged roster, the participants were alternately assigned to the two classes with an exception that, when two teachers from a single school were to be assigned, they were assigned to different teachers without regard for the roster location of the second. This precaution was taken to take account of the possibility that each would learn certain things not impressed on the other and, by comparing their work after their return to their school, both would benefit. These exceptions arose because the policy had been followed in making assignments to conferences, of placing all of the teachers from any school in the same conference to forestall the occurrence of the same kind of interaction between training methods that was being encouraged between classroom groups.

Under this arrangement, the only systematic difference between the two groups was the nature of their activity during the first hour of each morning's work. The members of the first (control) conference began each day by observing a demonstration class taught by the conference director, Professor Beberman; following that class the director conducted a half-hour critique session of the class' recitation for the conference participants. The class observed in this way was composed of ten volunteer pupils who had been selected for admission to University High School that September. These ten pupils resembled the class seen

in the films in age, academic background, and scores on the tests used by the High School for selection of sub-freshmen. Following this session, the teachers dispersed to their classrooms to begin the day's study of the subject matter.

The members of the second (experimental) conference began their day by attending a double showing of one of the subject series. After a brief orientation to the nature of the film, in the form of a one page description of the subject matter and the time intervals covered, they watched the film itself. Following this showing, they were given a discussion guide which consisted of a set of questions pertaining to that film and the work it portrayed. These questions tended to emphasize and encourage observation of the behavior of the pupils and teacher in the film and to point toward the pedagogical implications of those observations. After these questions were discussed by the viewers, the film was shown again so that the various interpretations which had been uncovered in the course of the discussion could be re-examined by the discussants. Following this showing, a critique of the content of the film was conducted along lines similar to that held in connection with live demonstration classes.

Two unexpected qualitative differences between the demonstration classes became apparent immediately after the opening of the second conference. The behavior of the experimental group in the post-class discussions was a matter of surprise to everyone connected with the conferences; not only did the members of the film-trained group find a greater number of questions and comments to bring forth in the critique session, but it was clear that these indicated a generally higher level of comprehension of the principles which the conference sought to impart. No quantitative data or statistical interpretations are available to support this view, but the difference was so great that it could not be overlooked by staff members who were familiar with the usual nature of such sessions.

A number of explanations have been advanced for the existence of this difference, that the perspective afforded by the film gives the teacher-trainee a stronger feeling of participation than does the view from the back of the room, that the deletion of nonessentials gives the viewer a more immediate awareness of the relevant aspects of the demonstration, that the time-compressive aspect of the films provides the viewer with a broader range of material on which to base comments, etc. It is not possible to determine which one or ones of these lies at the basis of the change; it is sufficient to note that a large and readily noticeable improvement in discussion behavior results from the use of films for this purpose.

The second unforseen advantage discovered in the course of this conference is in the changed role of the teacher of the demonstration class. In the past, the teacher who conducted that class was also faced with a critique session after it had been dismissed. The double role, played in this case by Professor Beberman, is a demanding one since the teacher, legitimately tired after an hour's exacting work in teaching, is not likely to be in a mood or condition for the further

session. The use of films not only relieves him of the task of teaching for the first hour, but actually places him in a better position to discuss the events of the class because he has had better opportunity and occasion to give thoughtful consideration to the kinds of questions and comments that the discussion class is likely to raise.

In both conferences, optional study sessions were held each evening under the supervision of a UICSM staff teacher for the benefit of the trainees who found themselves in need of extra-class instruction in their study of the course. In the experimental conference, the opportunity was also provided in connection with these evening sessions for re-running any of the films shown up to that point, but few of the conference members exercised this privilege.

IV. Design of The Experiment

The primary concern in evaluating the usefulness of these films for their intended purpose is the rather formidable one of finding (or of creating) and applying a measure of teacher effectiveness. An adequate evaluation of effectiveness requires that the evaluator specify what functions are performed by a good teacher and find general agreement in distinguishing between good and poor, or between effective and ineffective teachers. This problem has been attacked, without noteable success, by many researchers in a variety of contexts but, until the teacher's task has been defined more explicitly than it has yet been, this condition is not likely to improve.

The selection of criteria for objective valuation of teachers was influenced, in this case, by the emphasis that the UICSM places on teaching skill. Mathematics is especially sensitive to the skill of the teacher and there is much to be gained (or lost) by devoting careful attention to developing skill in this area. A qualitative evaluation of teaching ability can be made reliably by an experienced observer but it is difficult for that observer to specify the characteristics of observable behavior on which he has based his evaluation; for this reason, the subjective judgment of teacher-observers constitutes a good criterion against which objective measures may be compared, but it is not an adequate basis on which to base an evaluation.

Three indices of teaching skill are applied here; subject-matter examinations to measure mathematical competence without which no amount of pedagogical training will be of value, measures of pupil achievement as an index of the effectiveness with which that knowledge is utilized in the classroom, and a comparison of the extent to which the two groups are aware of the problems that are likely to be encountered in teaching First Course.

The first of these, measures of the teachers' knowledge of the content of the course, is so obviously related to their training that it needs no justification. Most high school teachers, even though they hold undergraduate degrees in mathematics, are not familiar with this material because it has traditionally been reserved for graduate study. The reason for this tradition is not known; it is abundantly clear that ninth graders are capable of learning it, but with very few exceptions, their teachers have never encountered it. In certain respects, learning this subject is easier for the ninth grade pupil than for his teacher. The pupil approaches his study of First Course completely unprejudiced with respect to its content; he has no preconceived notions concerning the proper content of beginning algebra or the most appropriate manner for teaching it. The teacher, on the other hand, must not only learn the material, but must also reconcile the differences between it and the algebra with which he is familiar so that, as a result, he is apt to devote an appreciable portion of his time and efforts to a task which never becomes necessary to his pupils.

The inferential function of measurements of teachers' knowledge of mathematics is somewhat less precise than its determination. It is a basic postulate of all teacher training that subject-matter competence is vital, but no quantitative assessment of the relationship between teachers' and pupils' subject mastery has been made. The precision of this comparison would be increased manyfold if it were possible to equate a specified difference between two teachers' knowledges of subject matter to some specified difference in the degree of achievement to be expected of their pupils, but the present state of knowledge in this area is not adequate to such a purpose. It is possible to make reliable measures of the teacher's knowledge, however, and to compare the two groups in this respect, but only the expectation of higher levels of achievement from higher levels of teacher competence can be justified.

The use of pupil competence as an index of adequacy of teacher training is related to the overiding concern for teaching effectiveness that permeates all of the work of the UICSM.. Just as the text materials are required to combine mathematical soundness and teachability, so also must teacher training result not only in subject matter competence, but also the ability to impart that knowledge. It would be possible, if only teacher's examinations were to be used as criteria, to develop a training program that would turn out highly competent mathematicians who were ineffective in the classroom, but knowledge of mathematics is an ancillary necessity which, in itself, has no educational value. The function of a teacher is not only to know, but to share his knowledge, and until his pupils have demonstrated that this has been done, the time and effort devoted to that teacher's training are poorly spent.

The third index of the usefulness of the films was a comparison of the teacher's expectations of the classroom use of several topics and his experience in teaching them. This comparison takes the form of rank-difference correlations between his estimates of the difficulty of certain topics made before he has taught them, and again after completing that Unit with his class. This comparison was included as an indication of the extent to which the teacher has understood the nature of each Unit, considered as a whole, and has simultaneously considered its effect on his pupils and his task in carrying it to them. The subject matter examinations, both for teachers and for pupils, were such that they could be completed successfully by considering a single topic or idea in connection with each question, even though the relationships between these ideas was not well understood. When he was asked to predict the relative difficulties of a number of topics, the teacher was forced to take into account the entire unit and to compare individual topics with one another. Because his ranking was made in terms of the expected difficulty for his class, he was also encouraged to think in terms of his situation in teaching, as distinguished from his own difficulties in learning, the material.

The two groups of teachers were exposed to the same mathematical content under the same teachers, so there is no reason to expect differences between groups in the nature of their ideas of the relationships between and among the topics they rated. There is a reasonable expectation of difference between the groups in their expectations of their

own classroom experiences in teaching these topics, however, because one is extrapolating from the basis of the observation of all of the classroom events in a narrowly restricted portion of the course, while the other is interpolating from their observation of a number of crucial events selected from a much broader range of subject matter and classroom events.

Nature of the Measurements

Comparisons between the two groups with respect to their subject-matter competence are the simplest to make. During the period of his attendance at the training conference, every participant completed all of the published examinations covering First Course; this includes a twenty-four item test covering each of the four units, and another twenty-four item final covering all of the units. These five scores may be summed and the total regarded as a measure of his subject-matter knowledge; inasmuch as the reliability of the single quizzes is at least 0.8, the reliability of the total scores, considered as a test 120 items in length, is of the order of 0.95.

Measurement of pupil achievement is considerably more elaborate. More time is available for testing and the information is used as the basis for a different kind of inference. The comparison of teachers on the basis of only five scores is justified because the need is only for the denial of a source of systematic variance considered possible but not at all probable. Because there was no a priori reason to suspect that a difference should exist, extensive disconfirmation was not required. In measuring subject matter knowledge of the pupils however, a distinct possibility of difference is present and a broader and more detailed basis from which to compare is required. Not only must the potential difference be allowed ample opportunity to appear, but provision must be made for its further investigation if it does appear.

It cannot be assumed that any collection of 120 test items is an adequate sampling of the content of a year's study so, in order to meet the demand for more comprehensive coverage of this material, a set of quizzes was prepared to probe, in greater detail, the extent of the pupils' knowledge. Eighteen quizzes, six for each of the first Three Units of First Course, were prepared in such a way that each represented approximately 15-20 pages (excluding review sections and discovery exercises) of text material. In this way, any difference that might develop between the groups could be located with sufficient accuracy to permit a reasonable inference concerning its nature. To as great an extent as was possible, the content of a single examination was confined to matters dealt with in a single film so that, should a measurable difference be manifested, it could be localized with respect both to the subject matter involved and to a specific aspect of the teacher's training. The content of these quizzes, designated by letters, is shown in the listing below (Table 6).

No quizzes were prepared in this series for the material of Unit 4 because the films by which these teachers were trained dealt almost exclusively with the first three units, so such quizzes would not be applicable to a specific evaluation purpose. Manny classes are not able to complete more than two units of work in a single school year

Table 6

Content of Criterion Quizzes

A. Use and mention. "Trip" interpretation of real numbers .

B. Adding real numbers.

C. Multiplying Real Numbers.

D. Using the distributive principle; identifying principles, instances and consequences.

E. Operations and their inverses, I; principles for real numbers.

F. Opposites and subtraction.

G. Comparing real numbers; open sentences, pronumerals, and substitutions.

H. Identifying principles in generalizations; basic simplifications.

I. Simplifying formulas; equivalent expressions I.

J. Equivalent expressions II; using principles and theorems as justifications.

K. Operations and their inverses, II.

L. Equivalent expressions, III; fractions and division.

M. Solution sets and the number-line.

N. Subsets; solution sets.

O. Transforming and solving equations.

P. Solving worded problems.

Q. Factoring and solving certain quadratic expressions and equations.

R. Solving inequations; square roots and decimal approximations.

and almost none go beyond three units in the first year so, even if such quizzes were used, they would yield only a small amount of data, and this from highly atypical groups.

All of the eighteen quizzes used in this experiment were prepared by Mr. Ronald D. Szoke, a staff teacher of the UICSM project.

The data pertaining to teachers' expectations and discoveries of topic difficulty are the simplest of all in their form. For each Unit, a list of ten topics was prepared and submitted to the trainee teachers at the conclusion of that unit during the summer conference, with instructions to rank those topics according to the difficulty they expected to encounter in teaching them. When each teacher reported that his class had completed a Unit, another questionnaire was sent to him with instructions to rank the topics in the order of the ease he encountered in teaching them; this questionnaire contained the same topics, of course, in a different order. The comparisons of interest in interpreting this material are the consistency of rankings within each conference group and the pairs of ranks assigned by each teacher.

Form of the Analysis

Data obtained from the teachers during their study of First Course are easiest to analyze because they are quite simple in form and because they do not, by their nature, require intensive investigation. The first step is that of justifying the summing of five scores obtained from each teacher; this is accomplished by a one-way analysis of variance which tests the hypothesis that all of these score distributions are characterized by comparable means and variances. The summed scores may then be compared by either of two methods (both are reported below); a simple t test is the most direct method of analysis, but they may also be treated as a two-dimensional array in which the two conferences constitute the rows, and the five quizzes the columns. This mode of analysis, although somewhat more involved than the direct comparison by means of t, yields more information since it allows simultaneous comparison of quizzes and conferences as well as the investigation of any interaction that may exist between them. Summary tables for this, as well as for all of the analysis described, are presented in a summation of results section.

The analyses conducted of quiz data obtained from the pupils are more elaborate and of greater length because they are far more abundant, were obtained under different conditions, and were used to justify a different sort of inference. Several alternatives are available in analyzing and interpreting this mass of data, two of these are used here.

The easiest and most obvious analysis would be to consider the entire group of pupils who completed these quizzes as consisting of two sub-groups, each made up of the approximately 1250 pupils who studied under teachers trained at one or the other of the two conferences. The distribution of scores from each quiz could then be subjected to a one-way analysis of variance as a test of the hypothesis of no difference. This is the equivalent to conducting as many separate

experiments as there are quizzes in the series; each experiment being of the type usually referred to as simple randomized. Such a procedure would be satisfactory if a new set of subjects were available for each replication or if the available subjects could be reassigned, by some random procedure, to experimental groups. Both of these procedures are obviously unfeasible so the continued replication of the comparison, using the same subjects grouped in the same way each time would soon have ceased to yield sufficient information to justify its continued use.

The lack of efficiency of this form of analysis, as well as the procedures by which it may be partially obviated, may be made clear by a brief examination of some characteristics of experiments which rest on variance analysis and the sources of variance with which such an analysis deals.

All experiments of this sort involve a quantitative observation of some aspect of behavior; in this case, the behavior observed is that of finding correct answers to questions based on certain mathematical content. These, like all measures, are never perfectly accurate, so the mean of a number of them is obtained and considered to be a more accurate measure than is any single observation. One of the most important characteristics of any set of measurements is the extent to which they are clustered around or dispersed from their mean. The variance in a set of measurements is a quantitative expression of the degree of this kind of dispersion. The failure of individual measures to fall at the mean of the entire distribution is attributed to two kinds of influences: systematic ones which are related to the operation of the experimental variable or to differences between individuals, and extraneous ones which are due to the operation of other influences which, troublesome and unavoidable as they are, are inevitably present in the experimental situation. The task of an experimenter is to conduct his data-gathering and analysis in such a way that the variables which give rise to the first kind of differences may be seen clearly, while those which give rise to the second are not permitted to bias the conclusion he draws from the experiment.

The effects of the extraneous circumstances present in an experimental situation are referred to as 'errors', and may be further classified into three categories. Errors based on the selection of subjects (Type S) occur when, through an accident of randomization, the levels of ability represented in experimental groups is not comparable. Errors based on group functioning (Type G) occur when a circumstance which has an effect, not related to the action of an experimental variable, on all of the members on one treatment group and on none of the members of the other. Errors associated with unforseen interactions between an experimentally constituted group and a treatment category (Type R) occur when that interaction is specific to that group and would not be likely to reoccur if the experiment were repeated with a new set of groups.

The test of significance applied to such a body of data is a comparison between the amounts of systematic variance, that associated with the experimental variable, and that which is associated with

inaccuracies in measurement. One of the factors of which account must be taken in an appropriate test of significance is the occurrence of one or more of the errors of the kinds just mentioned. If these cannot be eliminated from the experiment, the next best course is to assess their probable importance. Various kinds of significance tests are available, and the choice rests on considerations such as the selection of subjects, the manner in which they are assigned to treatment groups, the kind of criterion measures employed, and the mode of analysis contemplated.

The precision with which such a test of significance operates, i.e., estimates the true probability of a chance occurrence of a difference between groups, is governed by the correspondence between the categories of error that are present in the data and the categories taken into account by that test. If analysis is conducted in such a way that errors in the data are not taken into consideration by the test of significance, then the estimate made of the magnitude of the error variance will be smaller than it would be, if those errors were allowed to inflate the estimate, and that of the true (systematic) variance will be correspondingly larger. The inference made from such a comparison will be in error because the ratio true/error will be artificially inflated and the effects of the treatments will appear to be stronger than they actually are. On the other hand, an analysis that takes into account errors that are not present in the data will result in an overestimate of the error component, and a corresponding underestimate of the true component, with the result that the treatment effects will be made to appear weaker than they really are. It is a cardinal principle of experimental design that the form of the analysis be chosen in the light of a careful and realistic appraisal of the nature of the measurements because an over-precise significance test may vitiate a conclusion to as great an extent as an imprecise one.

In the analysis of the scores of teacher-trainees on their Unit examinations, a simple randomized design was appropriate enough because this form of analysis takes into account only errors of Type S. If G or R errors were present, the increase in total variance brought about by the artificial increase in one group's scores would not be accompanied by a corresponding increase in the estimate made of the error variance, and the ratio between them would be increased. If, for example, one of the groups had inadvertently been allowed to work for a few minutes longer than the other on one of their examinations (a simple and very obvious Type G error), then their scores would have been increased. This artificial increase would increase the amount of total variance, but would not have appeared in the estimate of error. The consequence would have been an artificially inflated significance measure to support a conclusion that the groups differed in their mastery of the subject-matter.

Not all of the possible sources of error are as easy to detect as this example, however, and a careful examination of possible errors is required before selecting a mode of analysis. The experiment described here is an unusual one in the manner in which these errors of measurement are found and, on that account, in the selection of the

kind of analysis to be made of the data. Type S errors are surely present in the data because the pupils from whom they were obtained differ in their subject-matter mastery, but to the extent that pupils randomly assigned to treatment-groups (through the random assignment of teachers to conferences), a test of significance which makes account of Type S errors will compensate for these. Differences in teacher ability, classified as Type S errors in the analysis of the teachers' quiz data, are Type G errors in this analysis because any teacher characteristic influences an entire class. There are other Type G errors, of course; the quizzes were teacher administered so it is probable that small differences existed between the classes in the manner of administration and these would affect the entire class' performance.

Type R errors may have sprung from sources such as systematic differences between groups in characteristics of the schools in which the program is used. Differences in admission policies, homogeneity of grouping, social or economic characteristics of the community, quality of the elementary schools from which the pupils were drawn, or the existence of other activities which influence the amount of time devoted by the pupils to their mathematics assignments all would influence their achievement. The only kinds of experiments which would make possible an estimate of the magnitude of such errors would contain other features which make it completely unfeasible. Two classes could have been scheduled in each school, one taught by a teacher from each experimental group to remove any bias from this direction, but this practice would have introduced a more serious source of contamination in the interaction between the two teachers and the sharing of their experience and knowledge. The other alternative would have been one or more replications of the entire experiment, including two more training conferences, and another year of data gathering. The cost of such a replication in time, money, administrative inconvenience, and delay in completion exceeds the probable cost of prejudicing the experiment by such an amount that this alternative was never given serious consideration.

The only remaining condition of the experiment which must be accounted for in the analysis is the variation in class sizes. If every class were of the same size, it would be sufficient for analytical purposes to use the sum of the pupils' scores as a measure of competence for that class for any quiz. Not only are there size differences between classes, but there are also differences within any class between quizzes due to absences, drop outs, transfers, etc., and instances in which a teacher neglected to administer or to report the results of a quiz for his class. The use of the sum of scores as a unit of analysis under these conditions would have the effect of ascribing greater weight to larger classes, since they would be more heavily represented in such a sum. The first of these effects would probably not constitute a large error, since it can be shown that in these data, the correlation between class size and mean score is negligible. The second circumstance is more serious because it would have the effect of ascribing greater weight to those quizzes for which the most data was available.

The difficulty associated with differences in class size may be avoided by using the class, rather than the pupil, as the unit of analysis. Instead of a group of individual scores, reflecting differences among pupils, an analysis may be conducted of the unweighted means of the classes, regardless of their size. The variance arising from individual differences is submerged in the class means, and the inferences drawn from such an analysis all pertain to the entire class. Subject differences, in terms of differences between pupils disappear in this case, but are replaced by another kind of Type S errors: extraneous circumstances affecting an entire class will be reflected in the mean so that an error such as deviation from standardized conditions in test administration, is removed from the C and into the S category. Type C errors may still be present from circumstances which influenced the teachers in one conference only; for example, a question may occur in a quiz which was discussed in a class session during one of the training conferences, or one of the teachers in the summer conference may have treated some topic at greater length in one of the sessions than in the other. Errors of the Type R are not changed by the change in the unit of analysis.

Difficulties associated with missing classes may be circumvented in either of two ways: if the quizzes are analyzed separately, differences in the number of classes reporting are accounted for by differences in the number of degrees of freedom assigned to each, or those classes for which data is missing may be omitted from the analysis. Each of these procedures is accompanied both by advantages and disadvantages; both are utilized and the results of each analysis will be reported.

In analyzing the twenty-one quizzes separately, it must be borne in mind that Type G errors may exist which are not represented in the significance test; to the extent that this occurs, the effects of the treatments (conferences) will appear to be stronger than they really are. There is no a priori reason to assume that such errors will occur with greater frequency or severity in one group than in the other, but neither is there any means by which such an occurrence might be detected in this analysis. For this reason, the *F* ratios reported in connection with the separate analysis of the quizzes should be regarded with suspicion. Another qualification must be introduced into the interpretation of an analysis performed as this one was; the stipulation of a five percent level of significance indicates that the *F* ratio which will be accepted as indicative of a difference between groups would be expected to occur once in twenty comparisons, even though no real difference existed. When twenty-one comparisons are made, as they are in this case, the existence of one such significant *F* can be attributed to the nature of the analysis rather than to the effects of the experimental variable. With these two qualifications in mind, the results of the analysis of the separate quizzes must be examined without expressing great concern for the existence of a single test which shows significance at the five percent level, or for the existence of significance levels near, but not yet at, that level.

The alternative to such a quiz-by-quiz analysis is one which, at the expense of discarding some of the data, eliminates the shortcomings of the randomized design discussed in the preceding paragraph. This experiment, referred to as 'Groups within Treatments' is applicable to situations in which the populations consists of intact sub-groups over whose composition the experimenter has no control. It is less precise for the number of subjects involved and requires, in this application, that all data be discarded for those classes which have not reported every quiz, but its advantages outweigh this lack of precision.

One such advantage is its applicability to the case in which it is important to avoid inter-group contamination. Because the class, rather than the individual is the unit of analysis, it is of less importance that all of the pupils in a class are subject to the same extraneous influence than that none of them is subject to any other influence. If such extraneous factors operate systematically in favor of one or the other of the treatments the conclusion is not invalidated because these same factors would probably operate in the same way if that treatment were widely used in practice; if, however, the groups were to be exposed to teachers trained in both methods, or if the teacher were to have access to the training offered in the conference which he did not attend, the analysis would be obscured and the inference invalidated.

Topic-Difficulty Rankings

In analysing the ranking made by the teacher-trainees of the difficulty expected in teaching certain topics from each unit, two points of view may be adopted. Every teacher in both conferences ranked a set of topics for each of the first three units, so a measure may be taken of the degree of agreement among teachers in each group, both before and after having taught it. A Coefficient of Concordance (Kendall's W) is used for this purpose and the interest in interpreting these centers, first on the differences between the groups in their expectations concerning the teachability of these topics, and second, on whether these differences are altered by having taught the material.

The second approach that may be made to this material is the examination of changes in attitudes of individual teachers toward these topics. For this purpose, a rank-difference correlation is computed for every teacher who returned a questionnaire after teaching the topics mentioned. Close agreement between expectation and realisation is interpreted as evidence for the applicability of the training method involved.

The Teacher's Quizzes

The first step to be taken in analyzing the accomplishments of the teacher-trainees in their mastery of the content of First Course in that of justifying the summing of their five quiz scores. This was accomplished by a one-way analysis of variance to determine whether the five sets of scores were comparable. The F ratio associated with this comparison is 2.15; that required for significance at the .05 level would be 5.63; this failure even to approach significance indicates that summing these scores and using the sum as a single measure is justified. With the summed scores available, a simple t test between the means of the two distributions makes it clear ($t = 0.3\overline{3}$) that any difference between the groups of teachers may be attributed to chance factors.

Another test of the same hypothesis may be made on an examination of individual quiz scores rather than on their sums by treating the data as a two dimensional array in which the two conference groups constitute the rows, and the five quizzes the columns. This analysis is summarized below in Table 7.

Table 7
Summary of Analysis of
Teachers' Quiz Scores

Source of Variance	df	Mean Square	F	F Required (.05)
Units	4	23.25	2.07	(4/345: 2.39)
Conferences	1	7.00	0.62	(345/1: 2.54.)
U x C	4	2.75	0.24	(345/4: 5.64)
Within groups	345	11.24		
Total	354			

The Units, as sources of variance, approach a significant level, suggesting the existence of a general difference between them in difficulty level common to all the teachers. This difference, if it exists, is specific to the quizzes involved rather than to the subjects, and little importance is attached to it. The very low F ratio between conferences confirms for five separate quizzes, the conclusion that the t ratio yielded for the single score; the small ratio for interaction indicates that the groups did not change their positions relative to one another during the course of their training. No difference, as large as that which can be attributed to the effects of random sampling, exists: (1) among the five quizzes for all teachers, (2) between the two groups for all of their quizzes, or (3) between the two groups on any one quiz.

On the basis of either of these analyses it can be stipulated that, insofar as their knowledge of First Course is concerned, the two groups of teachers were comparably trained and, by extension of that stipulation, that any differences which may exist among the achievements of their classes must be attributed to sources other than differences in the subject-matter competence of the teachers.

The Pupils' Quizzes

The first analysis made of the data from quizzes administered to the pupils was a set of individual comparisons of the distributions of mean scores from classes taught by teachers in the first conference with distributions of mean scores from classes taught by teachers trained in the second conference. The twenty-one pairs of means and the F ratios associated with each comparison are shown in Table 8. All of the quizzes from all classes are included in this comparison. The entries under the headings 1 and 2 in this table are mean scores for all of the class means from teachers attending that conference. The column headed 'Difference' shows the difference between the two sets of means in the direction, $C_1 - C_2$ (Control - Experimental), so that a negative difference indicates a higher mean score among the classes taught by teachers trained in the Experimental conference. The columns headed 'k' refer to the number of classes included in each mean, and that headed 'F' shows the ratio of the two variances. The one headed 'p' shows the probability of chance occurrence of a difference as large as the one which it accompanies.

Table 8

Means and Standard Deviations of Class Scores and Tests of Significance Between Them

Quiz	M_1	M_2	σ_1	σ_2	k_1	k_2	Diff.	F	P
A	13.41	13.96	1.15	2.08	43	46	- .55	2.38	.2 > p > .1
B	15.36	15.36	1.38	1.40	45	46	0	0	
C	15.43	15.13	1.46	1.93	44	46	.30	.65	
D	16.16	15.73	2.06	2.74	44	46	.43	.88	
E	13.22	13.12	1.74	2.58	45	46	.10	0	
F	11.99	11.82	1.73	2.03	45	46	.17	0	
U I	17.40	16.32	2.20	3.27	45	46	1.08	3.26	.1 > p > .05
G	12.92	12.46	1.70	2.12	45	43	.46	1.34	
H	14.48	13.46	1.73	1.99	45	44	1.02	6.78	.01
I	11.80	11.36	1.69	1.88	44	44	.44	1.22	
J	11.08	10.80	2.04	2.73	44	45	.28	.34	
K	9.85	9.87	1.52	2.00	45	46	- .02	0	
L	8.78	8.79	1.65	1.95	44	45	- .01	0	
U II	14.72	14.52	2.51	3.28	45	46	.20	.11	
M	10.36	10.61	2.30	2.43	44	42	- .25	0	
N	9.61	9.89	1.68	1.78	43	40	- .28	.65	
O	11.21	11.81	1.24	1.47	30	37	- .60	2.58	.2 > p > .1
P	8.21	7.92	1.52	1.08	26	30	.29	1.14	
Q	9.01	8.29	2.00	1.09	16	26	- .72	2.10	.2 > p > .1
R	7.28	6.63	2.13	1.65	14	15	.65	.51	
U III	14.96	13.53	3.17	2.44	15	23	1.43	2.49	

Examination of Table 8 reveals a number of interesting circumstances. Most of the differences between means are not large, but in fourteen of the twenty-one comparisons the classes of control conference teachers display a higher mean score than do classes of the film-trained teachers; the probability of chance superiority of one group in fourteen of twenty-one comparisons, when both are selected randomly from a single population is approximately 0.2, which is not compelling evidence for a difference between the groups.

Only one of the differences, that between the means of Test H, is significant beyond the .05 level and that one is well beyond. A single difference significant at that level may be expected to occur once in twenty comparisons without inferential significance, and nothing else is to be found in Unit II which even approaches significance, so the importance to be attached to this difference is questionable. Quiz H devotes ten questions each to identification of principles in generalizations and to basic simplification (pages 2-28 to 2-59 in the text); examination of the item statistics, which were tabulated separately for the two groups, shows that the difference lies in the questions of the second portion. This topic, basic simplification, is not an easy one to teach under any conditions, and it is especially taxing to the ingenuity and patience when taught by the method in which these teachers were trained. It is not uncommon to find teachers who, at this point in the course, temporarily abandon the discovery principle and instruct their classes in the traditional manner in which they are experienced. It may be hypothesized that the film-trained teachers, having observed more examples of it, were more acutely aware of the discovery principle and more deeply impressed with its usefulness than were the control teachers, and that they persisted in their attempts to apply it to this material, even though such attempts are not always rewarding to the teacher whose skill at discovery teaching is not yet developed. The control teachers, under this hypothesis, would have been more ready to abandon the attempt at discovery in favor of the traditional method which they knew, from their own experience, would make it possible for their pupils to learn to manipulate symbols with sufficient facility to show superiority in a short quiz.

The distribution of other differences among the units is difficult to interpret. Quiz A, for example, is the only member of the series dealing with material that was demonstrated to the control teachers but was not included in any of the films. It would be reasonable to expect that the pupils of these teachers would excel in this quiz, but the difference between the means of the two groups is in the opposite direction. The pupils of the teachers who had no specific training in this subject matter displayed, as a group, a higher level of competence with that quiz. This circumstance suggests (and other evidence to be reported later confirms the suggestion) that specificity in observational training is not always to be regarded as advantageous. It is possible that, because they were shown in considerable detail and at great length, one of the ways in which a topic may be taught successfuly, many teachers attempted to adopt that same method in their own classes, imitating the demonstration teacher without regard for the differences that exist between the classes, or for the fact that in their own knowledge of the subject matter, skill at this kind of teaching, and personal characteristics of temperament, they do not resemble the demonstration teacher.

The differences found in Unit III quizzes can be accounted for as artifacts related to the manner in which the experiment was conducted rather than the action of the experimental variable. Even though three relatively large differences are found in only seven comparisons, any assumed significance is dispelled when the nature and direction of those differences are examined. One of them (Quiz O) is in the direction of superiority of the experimental group, which makes its strongest showing in this unit. The difference in the Unit III Final is obscured by the presence of twelve classes, three from the control group and nine from the experimental group, for whom this test is included although they had not reached the end of the Unit. This circumstance is probably related to the appearance of bimodality in the Unit III Final results (See Appendix C).

Reference to the columns 'k' (number of classes included in each comparison) helps to clarify the unexpected behavior of the signs of the differences in this unit. Most of the forty-five classes taught by first conference teachers progressed through the text at a fairly uniform rate; at the end of the school year, many of them were at approximately the same point in the course, and the number of classes reporting drops sharply after Quiz N, indicating that fifteen classes stopped in the vicinity of page 3-30 of the text. The classes taught by second conference teachers demonstrated greater variance, both in speed and achievement. This can be clearly seen if the standard deviations of the two groups are compared; throughout Units I and II these are larger for the experimental groups' classes. By year's end the classes taught by second conference teachers were distributed over a much longer range of subject matter coverage, so the decline in the number of classes reporting, which begins with quiz M, is much less precipitous for this group. The first classes to disappear from the 'k' column, i.e., the ones which had covered the least material, were those which had also displayed the lowest level of subject-matter competence. The four least capable classes in the entire group were taught by teachers trained in the second conference; when these classes did not report quiz results past the Unit II final, the mean of the film conference teachers' classes, freed of that burden, began to rise and the standard deviation of that group grew smaller. This trend continued, with superiority being shown on quizzes M, N, and O by the experimental group, until all but the most advanced classes in each group have disappeared from the comparisons.

Because of this circumstance, less importance is attached to the data of Unit III than to the other Units.

An alternative method of analysis of the pupils' quiz data is possible in which, instead of treating the quizzes separately, simultaneous comparisons are drawn between conferences and quizzes. This is the groups-within-treatments previously described. In making this analysis, differences between quizzes are included in order to make possible a test of the possibility of interaction between quizzes and treatment groups. No particular concern is displayed for the quiz differences, as such, because the quizzes were not pretested or adjusted for equal difficulty levels, so there is no reason to expect them to be comparable in this respect. The analysis reported here was based, for this and for other reasons, on a set of means which have been adjusted so that the mean of each quiz, for both treatment groups, falls at an arbitrarily selected value of 100, and the standard deviation of each is twenty score points. The amount of total

variance is changed by this standardization, but always in the same amount for each group so the relative proportions of their contributions is unaltered.

Table 9 shows a summary of this analysis of the sums of fourteen adjusted mean scores of forty-one classes representing each conference group. The fourteen scores summed are the seven tests (six quizzes and the Unit Final) from each of the first two units; the forty-one classes are those from which all of these scores were available. The analysis is limited to the first two units because this is the most advanced point in the course that was reached by all of the classes and the amount of available data drops off sharply after the completion of Unit II (see the columns headed 'k' in Table 8). To have included more quizzes would not only have reduced the number of classes that could be included, but would have based the analysis on a selected and atypical group. The fact that the same number of classes are included from each group is a fortunate circumstance; fortunate because, although it is not a necessary condition for such analysis, it expedites the computation, and circumstantial because the requirement of the analysis for complete data is not related at all to the circumstances which caused some of the teachers from each group to neglect to administer or to report the results of one or more quizzes.

Table 9

Analysis of Sums of Derived Scores From 14 Tests (2 Units) for 41 Classes in Each Conference for Which Complete Data Was Available

Source Of Variance	Mean Square	F	Minimum Significant F (.05)
Tests	24	.09	2.21
Conferences	308	1.11	3.85
Interaction	299	1.08	1.70
Within Treatments	278		

This analysis makes it clear that the difference between conference groups with respect to the first two Units of First Course is quite small in comparison to the differences within the groups. The failure of the F pertaining to conferences even to approach the value required for significance indicates that, if two sets of forty-one classes each were selected randomly from a single homogeneous population, a difference between them as large as this one or larger, could be expected to occur. From this it is inferred that, insofar as its effect on pupils' performance is concerned, the kind of observational training given to the teachers is immaterial. The failure of the F pertaining to the conferences × quizzes interaction to achieve significance indicates that the two conferences hold the some position relative to one another for both units and, by inference, that either training method is equally efficacious for both units. Had one conference group shown superiority in Unit I and the other been superior by approximately the same amount in Unit II, the comparison between groups would have shown a negligible difference because the differences would have canceled one another, but the interaction variance would have been relatively large.

Topic-Difficulty Rankings

When the teachers ranked the ten topics presented to them in questionnaire form at the end of their study of each unit, not only were they making predictions of the difficulty they expected to encounter in teaching those topics, but they were also making an implicit statement of their own attitudes toward the material and of their perceptions of it. Consistency among raters on this occasion is interpreted as evidence that their training has given them a more uniform set of ideas about the task which confronts them and of the manner in which they propose to go about it.

When those teachers ranked the same topics again after their classes had completed that Unit, they appraised both their experience and their preparation for it. Inter-rater agreement in these rankings is evidence of comparability of experience and if one of two groups shows greater consistency in this respect than does the other, it can be inferred that their training for their task was more uniform and that they all profited by their experience in the same way. Inter-trial (intra-rater) agreement is interpreted as evidence of adequacy of training since the individual whose expectations coincide closely with experience is obviously well prepared for teaching that topic. If the teachers of one conference are able to agree with themselves more closely than are those trained in the other conference, this circumstance is interpreted as evidence of a difference between the two training methods in their comprehensiveness and realism.

The first step taken in the examination of these rankings was the calculation of a Coefficient of Concordance for each unit to test the extent of inter-rater agreement within each conference group. This coefficient is a comparison of the amounts of agreement that exist between raters and between topics and is interpreted in much the same way as a correlation coefficient, except that it varies only from 0 (the case in which there is no agreement at all between raters) to 1 (which could occur only if every rater ranked the topics in the same order); this coefficient never assumes negative values.

Table 10

Coefficients of Concordance for First and Second Rankings of Ten Selected Topics From Each of Three Units

Unit	Conf. 1	Conf. 2
I	.30/.26	.25/.34
II	.47/.64	.63/.60
III	.37/	.37

Table 10 shows the coefficients for both conferences for each of the first three units. No data of this sort is included for Unit IV because none of these teachers was expected to cover more than three units in his first year, and because the films they saw did not include any examples from the fourth unit. In each entry, the value at the left of the diagonal represents the Coefficient of Concordance computed from the rankings made at the time that unit was completed in the training conference and before it was taught to a class; the entry following the diagonal represents the same statistic computed from the ratings made of the same topics by the same teachers after having taught that unit.

The most obvious feature of this table is the relatively high degree of agreement among the raters in both conferences in their attitudes toward the material of Unit II compared to their feelings about Unit I and Unit III. It would appear from this difference that the pedegogical implications of Unit II, as they are presented in these conferences, are more clearly organized than those relating to the other Units.

The change in inter-rater agreement as a result of experience is slight in the case of the topics of Unit I; the first conference teachers are no more uniform in their judgments as a result of their experience, and the film-trained teachers from whom a more realistic pre-experience appraisal was expected, are only slightly more so. No test for the significance of differences between these coefficients is available, so it cannot be appraised quantitatively, but the suggestion is present that film-trained teachers are affected in a different way by their experience.

In Unit II, not only do both groups show a better degree of agreement in their pre-experience expectations, but the first conference teachers become far more consistent in their rankings as a result of their experience. It appears from these data that, in preparing teachers for Unit II, classroom observation prepares the teachers well to profit from their first year's experience, but that film observation results in a more uniform and more realistic set of expectations and attitudes than does classroom observation. This outcome is not suprising since the second conference teachers had been shown many sequences from that unit, while the first conference based their expectations on extrapolation and conjecture. The implications for Unit III training are not clear since so few teachers finished this unit that a similar analysis was not attempted.

A more detailed examination of these teachers' attitudes toward the material they have learned is possible if the topics they ranked are examined individually. In Table 11, a summary of the rankings assigned to individual topics is shown. In these tables, the median rank is shown for each topic as an indicator of the group's central tendency, and the interquartile range as the index of their dispersion. In reading these table, the columns headed '1st' contain information derived from the rankings made after studying the material but before teaching it, the columns headed '2nd' pertain to ranks given after having taught the unit. The rows labeled 'C_1' refer to information obtained from the first (control) conference, and those labeled 'C_2' to the second (film-trained) conference.

Table 11.1

Summary of Difficulty Rankings Assigned to Ten Topics by Two Groups of Teachers Before and After Teaching Them

Unit I

Topic		Medians 1st	Medians 2nd	$Q_3 - Q_1$ 1st	$Q_3 - Q_1$ 2nd
The operation absolute valuing	C_1	5.4	5.5	4.3	4.1
	C_2	4.4	5.0	4.7	3.4
The associative principles	C_1	1.9	1.7	1.3	3.3
	C_2	3.2	2.6	3.5	2.8
Using the principles to justify short cuts	C_1	6.2	6.9	3.0	3.6
	C_2	7.3	8.1	3.0	3.0
Operations as sets of ordered pairs, and their inverses	C_1	8.0	6.5	3.4	3.2
	C_2	5.4	6.3	4.6	4.5
Showing that one of two numbers is the opposite of the other number	C_1	3.8	4.9	3.2	4.5
	C_2	3.9	4.4	4.5	5.1
Comparing real numbers	C_1	3.0	4.8	2.5	3.4
	C_2	2.9	3.8	3.2	2.9
The distributive principle	C_1	5.1	5.0	2.8	3.7
	C_2	5.3	4.6	2.7	3.3
Distinguishing between instances and consequences of the basic principles	C_1	9.2	9.0	1.2	1.9
	C_2	8.6	9.3	2.7	1.8
The operation opposite	C_1	4.8	3.7	3.2	3.5
	C_2	4.6	3.7	2.4	4.6
The conventions for omitting grouping symbols, and identification of the principal operators in an expression	C_1	8.2	6.4	2.5	4.0
	C_2	6.7	7.0	4.5	3.8

Table 11.2

Summary of Difficulty Rankings Assigned to Ten Topics by Two Groups of Teachers Before and After Teaching Them

Unit II

Topic		Medians 1st	Medians 2nd	$Q_3 - Q_1$ 1st	$Q_3 - Q_1$ 2nd
Testing patterns	C_1	7.2	7.1	3.6	3.7
	C_2	5.5	6.2	2.5	4.5
Generating pronumeral expressions from a pronumeral expression	C_1	5.2	7.5	2.3	4.0
	C_2	7.1	6.9	2.3	2.5
Substitution	C_1	2.8	2.8	1.1	2.2
	C_2	2.8	3.0	1.1	1.9
Discovering patterns of expression	C_1	6.9	6.4	3.1	3.7
	C_2	5.8	6.6	2.3	2.7
Theorems involving opposition	C_1	7.7	7.9	3.1	2.9
	C_2	7.8	7.6	2.3	4.0
Using frames	C_1	1.8	2.0	1.6	2.7
	C_2	1.7	1.8	1.1	1.7
Using pronumerals and quantifiers to state generalizations	C_1	6.9	5.4	3.4	2.7
	C_2	7.8	5.7	3.7	3.2
Omitting multiplication signs	C_1	1.6	1.3	1.6	1.0
	C_2	1.3	1.4	2.0	1.2
Using the division theorem	C_1	8.8	8.5	3.0	3.3
	C_2	9.2	9.7	1.8	2.1
Uniqueness and cancellation principles for addition and multiplication	C_1	5.9	6.0	3.2	3.8
	C_2	6.1	5.7	4.3	1.2

Table 11.3

Summary of Difficulty Rankings Assigned to Ten Topics by Two Groups of Teachers Before and After Teaching Them

Unit III

Topic		Median 1st	$Q_3 - Q_1$ 1st
Dividing and averaging method of computing square root	C_1	7.8	3.3
	C_2	7.6	4.4
Using brace-notation to name the solution set of a sentence	C_1	4.5	4.0
	C_2	3.9	4.6
Solving equations and inequations intuitively	C_1	6.3	3.4
	C_2	5.0	5.0
Equation transformation principles	C_1	4.5	4.6
	C_2	5.0	2.9
Expanding and factoring pronumeral expressions	C_1	5.5	4.2
	C_2	6.4	3.1
Graphing the locus of a sentence	C_1	4.7	3.0
	C_2	5.0	3.4
Using the 'sample answer' technique to obtain an equation for a word problem	C_1	5.8	3.4
	C_2	7.5	3.4
Using geometric language to name the locus of a sentence	C_1	4.6	3.2
	C_2	3.7	3.2
Using the exponent '2'	C_1	1.2	0.6
	C_2	1.1	0.6
Degrees of approximation	C_1	9.4	1.9
	C_2	9.3	1.9

Low ranks were assigned to topics expected (or found) to be easier to teach, so an upward shift in median values is based on discrepancies between anticipation and realization in the direction of greater-than-expected difficulty, while a downward shift indicates that the topic was found to be easier than it was expected to be.

A simple tally of such shifts shows that the film-trained teachers encountered more unpleasant suprises in this respect than did those who observed the live demonstration class in both Units. Comparison of inter-quartile ranges in the same way slows that the number of instances in interrater agreement increases is greater for the film-trained teachers than for the others. The films, as training media, are made to appear less desirable than classroom observation from the standpoint of the accuracy of prediction, but more desirable from the standpoint of the degree of preparation for profiting from experience.

The topics which show the largest changes between the first and second rankings may be compared to the nature of the training received by the raters with respect to that topic. The fourth topic shown under Unit I, 'Operations as sets of ordered pairs', is the best example of a pattern that is suggested throughout the table. This topic was dealt with specifically in Film V of the series, but was not included in the demonstration class observed by the control group. The former group found it to be somewhat more difficult than they expected, while the latter shows a median shift in the opposite direction so that the two groups are able to consider it about equally difficult after having taught it. Inter-rater agreement among the film trained teachers increases with experience, while it remains at about the same level in the other group. The same pattern may be observed in the rankings of the topic "Absolute Valuing" which is treated in Film XI; the film-trained teachers under-estimate the difficulty of teaching this topic, but can agree more closely about its difficulty after having taught it.

The same effect is shown, to a diminished extent, in the topic 'Using numerals and quantifiers to state generalizations" which is referred to, but not at length, in Film X. In this case, both groups of teachers found the topic to be easier than was expected, but both were also able to agree more closely after teaching it.

If a pattern exists in these comparisons, it shows that the film-trained teachers tend to underestimate the difficulty of teaching topics which are treated in detail in the films but, having had experience with them in class, show a better degree of inter-rater agreement concerning that difficulty. It may also be true that the existence of such an effect is related directly to the amount of explicit attention devoted to that topic in the teachers' training. In this case, the films would be considered as a useful, but somewhat deceptive training device, since they inculcate in the viewers an attitude that is consistent, but not wholly realistic. In combination with the experience gained in teaching, the use of films would seem to be preferred to the live class observation since, despite their initial misconceptions, these teachers profit more from experience than do the observers of live class demonstration.

Neither of the comparisons of the topic difficulty ranks has taken into account the differences between individual teachers in the accuracy of their perception and prediction of the nature of these topics. Rank-difference correlations were computed for the two sets of ranks supplied by each teacher and are reported in Table 12. A high correlation is evidence of close agreement between expectation and experience and is regarded as an indication that the teacher was more adequately trained, by virtue of his possession of a more realistic attitude and a higher degree of integration of knowledge, than was the teacher whose original conception of his teaching task were not in close accord with his classroom experience.

Table 12

Distributions of Rank-Difference Correlations between Expected and Experienced Difficulties of Teaching Ten Selected Topics from Two Units

	Unit I		Unit II	
	C_1	C_2	C_1	C_2
.86 - .95		3	6	1
.76 - .85	2	6	4	6
.66 - .75	3	4	4	7
.56 - .65	4	1	9	6
.46 - .55	3	4	2	2
.36 - .45	6	7	1	2
.26 - .35	4	5	2	
.16 - .25	6	4	2	1
.06 - .15	2	1		
-.06 - +.05	3			1
-.07 - -.16	2			
Median	.36	.46	.64	.66
$Q_3 - Q_1$	.38	.47	.25	.20

In Unit I, simple inspection suffices to show that the film-trained teachers found what they expected to a considerable greater extent than those trained by live class observation. This difference is interesting in comparison with the lack of inter-rater agreement shown above. The fact that some of these teachers were able to agree quite closely with their own initial ratings, while not agreeing with one another is not a contradiction if it is borne in mind that each teacher was asked to predict the difficulty for his own class. A teacher who knows that he is going to teach a relatively slow class could make a very realistic and valid appraisal of those topics which would not agree well with another teacher's equally accurate appraisal, if the second teacher were projecting his expectations to a highly competent class. The difference between groups is not as easy to explain; the first conference teachers' failure to agree with one another can be accounted for, but the existence of a relatively large proportion of teachers who fail to agree with themselves is interpreted as evidence that their training for teaching Unit I does not equip them for their task as well as film-training would have.

Other Analyses

In addition to that already reported, other kinds of information were obtained in connection with this study and were examined for whatever contribution they might make to understanding the processes involved in teaching and learning mathematics. Although they cannot be considered as a part of the formal evaluation of the film series, these contain some interest. Two of them represent unsuccessful attempts to relate external variables to the evaluation of teaching and two others give indication of containing useful relationships which, although they cannot be offered in support of the results reported above, facilitate their interpretation. These sub experiments are reported here, less for their contribution to the film evaluation than for their applicability to the planning of further research in his area.

The Attitude Scale

A scale of attitudes toward mathematics was constructed according to Thurstone's method and was applied to pupils at the beginning and again at the end of the school year, both in UICSM and in conventionally taught classes. The attitude changes were slight and so irregular that they seem random and it is assumed that the failure to show any systematic attitudinal shift is due to the lack of sensitivity or precision of the scale or the inapplicability of repeated administration of such an instrument at this age level.

About fifty statements pertaining to mathematics were written and submitted to about 600 eighth and ninth graders in mathematics classes in schools in Pekin, Illinois, to be evaluated in terms of the strength of the attitude that each expressed.

The conversion of these ratings into an equal-interval scale was complicated by two circumstances related to the nature of the subject population. One was the tendency of these pupils to regard all statements as indicating either extreme enthusiasm or undisguised antipathy toward mathematics; the admissibility of guarded or nearly-neutral statements of opinion is not recognized by many of these pupils. The result was a peculiarly shaped distribution of ratings of those statements intended to occupy scale positions near the neutral point; it was necessary to prepare a new set of statements and resubmit them, with revised instructions, to another set of raters. The second set was rated by approximately 600 pupils in the eighth and ninth grades of schools in North Plainfield, New Jersey. The second difficulty encountered in assigning scale values to these statements was the failure of many pupils to distinguish between judgments of an attitude expressed in these statements and their own reaction to the statements.

A suggestion appears in examination of these ratings that two criteria for judgment are being used; some of the statements seem to constitute a family which is characterized by implications that mathematics is interesting or pleasant, while others seem more to be concerned with a belief that mathematics is important or useful.

Twenty two statements were selected on the basis of low dispersions along the continuum of favorability and approximately equal spacing of their medians along the scale and these were submitted to several classes taught by conference trained teachers at the beginning of the year with instructions to indicate agreement or disagreement with each. About thirty of these classes were asked to restate their attitudes at the end of that year on the same form. The analysis of these results yielded no evidence that attitude shifts had taken place, either between administrations or between conference groups. The details of the analysis will not be reported here since there is so little evidence of any attitudinal changes, but the statements used and the scale values attached to each are shown in Appendix D.

In view of the difficulty encountered in finding statements which meet the conditions specified for inclusion in a Thurstone Scale, it is not surprising that the results of its application were not rewarding.

The Semantic Differential

A second attempt at providing an index of pupil's attitudes toward mathematics at a different level was made by examination of the factorial structure of ratings made by about two hundred pupils, of twelve mathematical concepts according to the method described by Osgood, et al[1]. In this procedure, a number of adjective pairs (antonyms) are selected and raters are asked to assign each concept being judged to a place along a seven-point scale whose ends are defined by one such pair of adjectives. The ratings are intercorrelated (between adjectives, summed across concepts) and the correlations are factor analyzed. It has been established that, over a wide range of concepts and adjective scales, three factors consistently appear. The strongest of these is defined by scales such as 'good-bad' and is labeled Evaluation; the second, defined by scales such as 'strong-weak' is labeled Potency; the third, defined by scales such as 'quick-slow' is labeled Activity.

It has been suggested that the loadings on the scales of the Evaluation factor may consititute a measure of opinions or attitudes toward the concepts in the population which provided the ratings. The use of the Semantic Differential in this context was based on this suggestion.

The twelve concepts judged in this analysis were:

prime number	multiplication	cancellation principle
inequations	graphs	associative principle
arithmetic	equations	$x^2 - y^2 = (x - y)(x + y)$
problems	zero	$(x + y) + 3 = x + (y + 3)$.

In selecting the scales along which these concepts were to be judged, three adjective pairs were included which were known to pertain to each of

---

[1] Osgood, C. E. Succi, G. J., and Tannenbaum, P. H., The Measurement of Meaning. Urbana, University of Illinois Press, 1957.

the three established factors, and three more were added which, although their factorial content was not established, were applicable to these concepts.* The scales used are shown in Table 13 which shows the factor loadings obtained from the analysis of the responses of approximately one hundred ninth graders enrolled in four UICSM courses in schools in Elkhart, Indiana.

Table 13

Rotated Factor Loadings
An Analysis of Ratings Made Of
Twelve Mathematical Concepts

	A	B	h²
Evaluation Scales			
Unpleasant-pleasant	-.96	.24	.96
Sweet-sour	.97	-.10	.93
Dark-bright	-.97	.11	.95
Activity Scales			
Complex-simple	-.97	-.01	.94
Slow-quick	-.96	.20	.97
Active-passive	.18	-.96	.95
Potency Scales			
Strong-weak	.01	-.97	.94
Light-heavy	.95	.22	.95
Soft-hard	.96	.16	.95
Experimental Scales			
Clear-hazy	.93	-.25	.89
Empty-full	.16	.97	.97
Relaxed-tense	.98	.03	.97

Similar results were obtained from the analysis of the ratings made by approximately the same number of pupils enrolled in non-UICSM classes in the same school except that some confusion on the part of these pupils with respect to terms like "associative principle" and "inequations", which they had never encountered, made the structure slightly less obvious.

The results of this analysis are quite clear and quite disappointing. The second factor, defined by the scales, Strong-weak, Full-empty and Active-passive appears to be a dynamism factor sometimes found when connotative scales are used or affective reactions are sought. The other factor, defined by the other nine scales, appears to be an "easy pleasantness combination. It is interesting to compare the possible meaning of these results with the suggestion of two kinds of judgments made by pupils

---

*The author acknowledges his gratitude to Professor Osgood for his advice and assistance in selecting these scales and in interpreting the results.

who provided the standardisation data from the Thurstone scale.

The unusually high communalities in this analysis result from the tendency of these subjects, like those in the attitude scale standardisation group, to ignore the middle categories of the scales and concentrate their ratings at the ends of the scales. The effect of this practice is to inflate the correlations obtained from the ratings because, in computing such correlations, the weight given each response is proportional to the product of its distances from the two neutral points.

It can only be concluded from these two questionnaires that moderation of opinion is not a common characteristic among eighth and ninth grade pupils, and that account must be taken of this proclivity for emphatic expression of all opinions in preparing questionnaires for use with such groups.

Control Tests

When a suspicion began to arise that there existed an appreciable difference between the pupils in the two groups, arrangements were made to administer a set of tests of primary abilities as a means of confirming or denying that suspicion. It is not possible, of course, to use information obtained under such circumstances to make a direct comparison between the groups, but these tests were able to cast interesting light on the nature of the difference between them.

Direct comparison by means of these tests, i.e., the use of such scores to support the specification that an initial difference in ability had existed between the groups, is prohibited because such a specification would have been based on data obtained from the same subjects with whose ability the experiment is concerned. The use of test scores for this purpose constitutes 'bootstrapping', the error committed when predictive (or postdictive) measures are applied to the sample from which they were derived. Under these conditions it is difficult not to show a relationship but, when another trial is conducted with a new sample, that relationship is almost certain to diminish and may disappear completely.

Five tests from each of two published batteries were tried with these subjects and four of the ten were retained for further use. The batteries were the Guilford-Zimmerman Aptitude Survey [1] and the Employee Aptitude Survey [2].

A Verbal Comprehension test was utilised from each. That from the G-Z Survey was found to be too difficult for this group and its use was discontinued; that from the EAS was retained and administered to nearly all of the pupils involved in the experiment.

---

[1] The Guilford-Zimmerman Aptitude Survey, Sheridan Publishing Co., Beverly Hills, California.

[2] The Employee Aptitude Survey, Psychological Services Inc., Los Angeles, California.

A test of arithmetic computation was also included from each of the batteries; both required the solution of a series of fairly simple problems based on the four basic operations and were highly speeded. Difficulty levels for both tests were suitable for the sample tested but that from the EAS was chosen for further use because, being printed on a single sheet, it is more economical and more convenient to handle than the other which is in the form of a folded booklet.

Both of the space perception tests from the Guilford-Zimmerman Survey were abandoned because they were too difficult for the pupils. The test, Spatial Orientation, requires judgments of the effect on the perceived environment of changes in the position of the observer; the test, Spatial Visualization, requires judgments by an external stationary observer of the appearance of an object after it has been rotated one or more times.

The tests from the Employee Aptitude Survey were originally developed for selection of industrial workers who apparently resemble the eighth and ninth graders in this sample more closely than do the college students and adults for whom the Aptitude Survey was intended. The Symbolic Reasoning test from the EAS requires the evaluation of a series of single propositions stated in terms of the symbols normally used by mathematicians and logicians. The Verbal Reasoning test consists of a series of verbally stated complex propositions. Both of these require the evaluation of stated conclusions as true, false, or uncertain. The test, Numerical Reasoning, from the EAS battery was abandoned when it was determined that it did not measure an ability that could not be accounted for by the four tests already in use. The results obtained from the administration of these tests are summarized in Table 14 below.

Table 14

Summary of Results From Four Tests of Primary Abilities in the Two Groups

Test	k_1	k_2	M_1	M_2	σ_1	σ_2
Verbal Comp.	43	43	42.99	42.62	6.27	7.88
Number Ability	37	36	42.75	42.02	7.10	7.64
Verbal Reasoning	43	42	32.05	31.16	3.43	5.57
Symbolic Reasoning	43	42	29.95	28.69	4.10	5.29

Comparison of the two conference groups with respect to these four abilities leads to the same conclusion that was drawn from the longitudinal examination of their algebra quiz scores. The pupils taught by teachers trained in the experimental group show a wider range of scores and a slightly but consistently lower mean score than those taught by teachers trained in the first conference. The same pattern was noted and commented on in the comparison of the quiz scores reported in Table 8 (see page 37). Its recurrence here, in a context which is not related to the skill or training of the teacher, lends support to the

explanation proposed there for the consistent superiority of the control classes in this experiment. The unforeseen circumstances involved in the assignment of teachers to groups and in the assignments of classes to these teachers which placed the four least able classes in the experimental group has introduced a systematic difference in group variances which makes a precise comparison impossible.

Correlations with the Control Tests

The fourth sub-analysis carried out on the data obtained in this experiment was a correlational study involving the four control tests and the subject-matter examinations. The results of this analysis, although they do not pertain to the film evaluation, are reported here as a matter of general interest.

After the quiz scores and control test scores had been converted to a standard distribution with common means and variances, the mean scores of approximately eighty classes for all four control tests and the sum of the means for seven subject matter tests were intercorrelated. The number of classes included in the correlations varies because not all of the control test information was available from every class. The analysis was confined to Unit I test scores because the three units are known to correlate highly with one another so that the results of this analysis may be considered as applying also to Units II and III.

It should be emphasized that these correlations, shown in Table 15 below, are based on class means rather than on individual pupils' scores.

Table 15

Correlations Between Control Test Scores and Unit I Totals

	U I	V C	N A	S R
Unit I total				
Verbal Comp.	.76			
N. Ability	.61	.62		
Symbolic Reas.	.62	.63	.16	
Verbal Reas.	.63	.56	.35	.37

Careful examination of these correlations will reveal that the principal components of the class' performances with the quizzes of Unit I are those contained in the Verbal Comprehension and Numerical Ability tests: Verbal and Symbolic Reasoning correlate approximately as well with one another and with Verbal Comprehension as with the Unit Score (criterion) examination, from which it appears that the verbal nature of these tests is providing the source of the covariance. The Numerical Ability quiz provides a small amount of unique variance, justifying its inclusion in any attempt that might be made at prediction in this way.

Summary

A series of twenty films was prepared from footage obtained during the course of 77 meetings of a normally scheduled class in the University of Illinois High School during the academic year 1959-60. Each of these films is a half-hour in length and is devoted to one topic from the UICSM First Course [beginning algebra]. By careful editing, each topic is followed through its inception and development, even though its occurrences in class may be widely separated in time.

The series is intended to replace the classroom observation which the UICSM considers to be a necessary part of its teacher training program. Two advantages of films for such observational training are postulated: the administrative problems connected with assembling and teaching a demonstration class are eliminated so that such training may be made more readily available to a greater number of teachers and, by concentrating each film on a single topic, the teaching strategies devised and utilized by the UICSM are made more apparent to the observer.

The films were used in training a group of 34 teachers at a summer conference held at the University of Illinois during the summer of 1960. Their attitudes and the performance of their pupils were compared to those of another group of teachers trained at another conference held that summer in which the observational training consisted of daily attendance for a period of four weeks at a demonstration class taught by an experienced teacher. This report describes the films, the manner in which they were made and used, and the outcomes of the comparisons made of these two groups of teachers.

The subject matter competence of the pupils taught by these two groups of teachers was measured by eighteen quizzes and three Unit Examinations administered throughout the school year. The results of this series of tests, totalling 433 items to which six hours of working time were allotted, were compared according to two separate rationales. The two comparisons agree in indicating that the pupils of the film-trained teachers represent a considerably wider range of ability than the pupils of those teachers whose observational training was given in a live classroom. This difference is large enough to make comparisons of the two groups difficult and to mask the difference between them. It is attributed to an unforeseeable and unfortunate circumstance in the assignment of teachers to conference groups which placed all four of the least capable classes in the entire comparison sample in the hands of teachers trained with the film observation conference. If the data from these classes are excluded from the comparisons, there is evidence of a substantial difference between the groups in only one of the twenty-one comparisons.

The results of a set of four tests of primary abilities agrees with the analysis of the subject matter tests in identifying the same four classes as being markedly poorer than the remainder of the group.

The applicability of the films to teacher training was assessed by

asking each teacher, as soon as he had completed his study of a unit in the training conference, to rank-order a set of topics from that unit with respect to the difficulty he expected to encounter in teaching them. The same topics were ranked again, after the class had completed that unit, with respect to the difficulty actually encountered. These ranks were compared to determine the extent of agreement within each set, and the degree of correspondence between expected and experienced difficulty.

Three conclusions were drawn from this comparison:

[1] In judging the difficulty of topics within Unit 2, both groups of teachers are in better agreement with one another than in judging the difficulty of topics within Units 1 and 3. The film-trained teachers are more uniform in their attitudes toward these topics than are the control group teachers, but this superiority does not extend to the other units. It is concluded that the films are more effective in dealing with Unit 2 than with the first and third units.

[2] Agreement between pre and post experience judgments is better for film-trained teachers, as a group, than for those who observed a live class. It is concluded that film-observation provides a more realistic basis for teachers' expectations than live class observation.

[3] Discrepancies between expectation and experienced difficulties among the film-trained teachers are more common with respect to topics which are treated explicitly in the films. It is concluded that the films are most efficient when they depict general features of the teaching procedure and that they tend to make specific topics appear deceptively easy. This shortcoming might be overcome by the inclusion in the series of a film which presents an unedited class session to indicate the true, as contrasted with the edited, course of the development of one of these topics.

A Thurstone attitude scale developed for this purpose failed to show significant shifts of attitude, either between groups, or between administrations within either group. The Evaluation scale from a Semantic Differential analysis, as an indicator of pupil's attitudes, was also tried and discarded.

Roster of Control Group Teachers

114 Biscoe, John
129 Weeks Jr. H. S.
144 Newton Centre, Mass.

139 Burke, S. M. Gabriella
Sacred Heart School
Newton Centre, Mass.

140 Cook, Miss Margaret M.
Redford H. S.
Detroit 19, Michigan

151 Corcoran, Francis X
F. A. Day Jr. H. S.
Newtonville 60, Mass.

*134 Dunham, Robert
Mount Everett Regional
Sheffield, Mass.

101 Fellows, Mac C.
Melbourne Sr. High
Melbourne, Fla.

130 Fenske, Clara
137 Steubenville H. S.
Steubenville, Ohio

142 Filochiffer, Sister Ellen Denisse
Cathedral Preparatory School
Erie, Pa.

123 Flanagin, Raymond J.
Walnut Jr. H.
Grand Island, Nebr.

113 Goosens, Robert L.
Elkhorn Jr. H. S.
Maybeury, West Virginia

107 Hintenlang, Joseph
Council Rock H. S.
Newtown, Pa.

108 Jones, Rodney
147 The Campus School
State University College of Education
New Paltz, New York

103 King, Lili
Southwest Jr. H. S.
Melbourne, Fla.

136 Lindquist, Lloyd
152 Wilbur Wright Jr. H.
Munster, Indiana

153 Macalouski, S. M. Cyrille
Fontbonne Academy
Milton, Mass.

126 McCready, Douglas
The Principia
St. Louis, Mo.

105 McNally, Rev. Patrick
Holy Ghost Preparatory
Cornwells Heights, Pa.

*111 Miron, Vincent
Bigelow Jr. H.S.
Newton, Mass.

112 Murphy, Gerard
George Washington School
Ridgewood, New Jersey

122 Murray, Sister M. Campion
Mt. St. Joseph Academy
Brighton, Mass.

125 Nelson James Conrad
131 Boulder City H. S.
Boulder City, Nevada

102 Nelson, Thomas H.
*146 York Community H. S.
Elmhurst, Illinois

128 O'Neil, Sister M. Adelbert
Matighon High School
Cambridge 40, Mass.

117 O'Neil, Sister M. DePaul
St. Thomas Aquinas High
Jamaica Plain, Mass.

* Data collected but not included in analysis.

109 Peguesse, Mrs. Frances
Southwest Jr. H. S.
Melbourne, Fla.

*127 Ployer, Mrs. Frieda
Meadowbrook Jr. H. S.
Newton Centre 59, Mass.

104 Prickett, Morris
Benjamin Franklin Jr. H.
Ridgewood, New Jersey

139 Saalfeld, Rev. Charles
Monroe H. S.
Fairbanks, Alaska

*119 Sperger, Herbert
Thomas Williams Jr. H. S.
Wyncote, Pa.

140 Steward, Fred R.
Council Roch H. S.
Newton Bucks Count, Pa.

134 Taylor, Nancy
Ottawa Jr. H. S.
St. Clair Shores, Mich.

132 Troutner, Arthur
Pueblo High School
Tucson, Arizona

106 Wilder, Martha
Southwest Jr. H. S.
Melbourne, Fla.

118 Yawin, Robert A.
E. O. Smith H. S.
Storrs, Conn.

135 Yingst, Dale
Grant Comm. H. S.
Fox Lake, Illinois

145 Zuehlke, Minnie A.
Walnut Jr. H. S.
Grand Island, Nebr.

110 Zwoyer, Russell
115 York Comm. H. S.
121 Elmhurst, Illinois
154

Roster of Experimental

Group Teachers

233 Allen, Charles
Brentwood H. S.
Brentwood, N. Y.

243 Annett, William
249 Seaford H. S.
Seaford, N. Y.

223 Assante, Joseph
Arthur L. Johnson Reg. H.S.
Clark, N. J.

201 Bogart, Erma
Wheatley School
Old Westbury, N. Y.

210 Buck, Richard
Shaker H. S.
Newtonville, N. Y.

205 Carlson, Sister M. Clarice, C.D.P.
St. Mary H.S.
Sharpsburg, Pa.

206 Carothers, Alfred
Desert Sun
Idyllwild, Calif.

228 Deeds, Samuel
Western Ill. Univ. Lab. School
Macomb, Ill.

** Delaney, Ralph D.
Wiley Jr. H. S.
University Hts., Ohio

219 Eisenmenger, Sister M. Petrona
St. Francis H. S.
Mishawaka, Ind.

208 Enkoff, Ralph L.
Lab School, Ind. State Teachers
Terre Haute, Ind.

214 Foley, Dorthy C.
Wheaton & Kingwood H. S.
Kingwood, W. Va.

204 Foster, Alan
Willowbrook H. S.
Villa Park, Ill.

211 Guerriero, Andrew
216 Pasack Valley Regional H. S.
247 Hillsdale, N. J.

203 Guiliani, Edward R.
207 Pasack Valley Regional
239 Hillsdale, N. J.
248

**217 Holstein, Milton
Calvert Hall
Baltimore 1, Maryland

222 Hungerman, Sister M. Fabian, O.S.F.
245 St. Benedict Academy
Pittsburgh 29, Pa.

**215 Jones, Ruby
Permian H. S.
Odessa, Texas

213 Koczan, Sister M. Lucy, V.S.C.
Vincentian H. S.
Pittsburgh, Pa.

245 Lauth, Sister Joan Marie, C.D.P.
St. Basil H. S.
Pittsburgh 10, Pa.

227 Musser, John
No. Plainfield H. S.
North Plainfield, N. J.

242 Neblett, Mary Sue
Dial Jr. H. S.
Pine Bluff, Arkansas

226 Ober, Stephen S.
229 N. Shore Country Day School
Winnetka, Illinois

231 Oldham, Donald
241 Cripus Attucks
Indianapolis, Ind.

** Data not collected from these classes.

220 Rooney, S. M. Ignatius Loyola, RSM
221 Canevin H. S.
Pittsburgh, Pa.

244 Rucker, Walter
Edward M. Cope Jr. H. S.
Redlands, Calif.

212 Salkin, Monroe
Brentwood H. S.
Brentwood, N. Y.

218 Schendel, Sister M. Urban, RSM
Immaculate Conception H.
Washington, Pa.

238 Schroer, Wilfred H.
Hixson Jr. H. S.
Webster Groves 19, Mo.

235 Snell, Ronald D.
240 Mariemont H. S.
Cincinnati 27, Ohio

246 Strein, George
Communith H. S.
Blue Island, Ill.

236 Subco, John
Chippewa Jr. H. S.
St. Clair Shores, Mich.

234 Toney, Lloyd W.
Hixson Jr. H. S.
Webster Groves 19, Mo.

237 Trenbley, Miss Mary Agnes
North Plainfield H. S.
North Plainfield, N. J.

224 Wheeler, Herbert
Wheatley School
Old Westbury, N. Y.

202 Whitaker, Miss Wilhelmina
Hopkins Twp. H. S.
Granville, Ill.

209 Wight, A. Theodore
230 Hingland High
Salt Lake City, Utah

Appendix B

Characteristics of Participating Teacher Groups

Experience (Years)	C_1	C_2
0 - 2	5	10
3 - 5	5	6
6 - 8	7	8
9 - 11	3	3
12 - 14	4	2
15 - 17	2	1
18 - 20	1	
21 - 23		2
24 - 26	2	
27 - 29	2	
30 - 32	1	
33 - 35	1	1
36 - 38		1
Not Known	5	3
Total	38	37

Enrollment	C_1	C_2
1 - 249	3	4
250 - 499	8	5
500 - 749	7	5
750 - 999	5	1
1000 - 1249	7	8
1250 - 1499	2	3
1500 - 1749	1	5
1750 - 1999		1
2000 - 2249	2	1
2250 - 2999		
3000 - 3249		1
3250 - 3499	1	
Not Known	2	3
Total	38	37

Age (Years)	C_1	C_2
20 - 22	3	1
23 - 25	2	2
26 - 28	3	9
29 - 31	3	5
32 - 34	7	1
35 - 37		5
38 - 40	4	1
41 - 43	3	4
44 - 46	4	4
47 - 49	1	1
50 - 52	3	1
53 - 55	2	1
56 - 58	2	1
59 - 61		
62 - 64	1	1
Total	38	37

Math Units	C_1	C_2
0 - 3	1	1
4 - 7		1
8 - 11	1	1
12 - 15	3	1
16 - 19		2
20 - 23	3	3
24 - 27	5	7
28 - 31	5	5
32 - 35	4	5
36 - 39	4	1
40 - 43	3	3
44 - 47	2	2
48 - 51	1	1
52 - 55		1
56 - 59	2	1
60 - 63		
64 - 67		2
68 - 71	2	
72 - 75		
76 - 79	1	
Not Known	1	
Total	38	37

Appendix C. 1
Score Distributions of Criterion Quizzes (proportions)
Unit I

	A_1	A_2	B_1	B_2	C_1	C_2	D_1	D_2	E_1	E_2	F_1	F_2	U_1	U_2
													033	036
24		*001					002	002					068	057
23		003					026	013					086	064
22		008					046	046					087	089
21	005	014					059	065						
20	006	003	046	039	065	073	082	082	005	006	004	003	099	084
19	030	052	078	091	105	114	118	103	035	038	009	014	106	070
18	042	053	129	110	139	120	106	094	066	077	038	017	091	080
17	060	077	139	149	159	136	094	088	083	101	043	041	074	072
16	094	098	159	138	109	108	083	089	102	084	059	074	060	065
15	119	110	123	141	098	091	075	071	100	079	102	085	056	057
14	122	124	103	109	082	079	065	070	115	107	109	095	054	049
13	144	111	071	082	065	063	065	053	097	097	102	116	043	048
12	130	097	052	048	041	045	048	046	088	093	125	102	030	042
11	096	076	032	034	032	032	034	048	184	074	101	112	033	037
10	067	064	022	021	027	037	031	035	077	068	079	090	019	032
9	040	044	015	012	030	031	025	031	059	065	068	068	021	037
8	031	026	005	003	020	029	017	018	039	044	057	064	019	025
7	014	009	009	007	009	024	010	017	022	031	045	057	008	022
6	002	004	006	009	012	009	006	012	018	025	028	026	007	010
5	002	002	005	003	003	006	003	011	004	007	017	016	001	011
4			004	002	002	002	005	003	004	004	010	013	003	010
3			002	002	002	002	001	003	001		003	002	001	002
2				001						001	001	002		002
1												001		
0														
N	1256	1276	1299	1250	1271	1254	1247	1252	1297	1252	1270	1228	1292	1247
Mean	13.4	14.1	15.4	15.4	15.4	15.1	16.2	15.8	13.3	13.2	12.1	11.8	17.5	16.4
S.D.	2.86	3.30	3.04	2.95	3.40	3.66	4.00	4.78	3.35	3.56	3.38	3.39	4.30	4.98

Appendix C. 2
Score Distributions of Criterion Quizzes (proportions)
Unit II

	G_1	G_2	H_1	H_2	I_1	I_2	J_1	J_2	K_1	K_2	L_1	L_2	U_1	U_2
24													014	013
23													026	023
22													024	035
21													044	058
20	015	012	037	025	009	004	007	007					056	047
19	029	033	082	057	016	021	015	021	004	008			056	063
18	040	044	103	093	020	017	032	038	004	008	001	001	076	070
17	079	055	114	073	041	034	043	043	010	014	004	004	067	062
16	074	078	119	112	065	056	055	052	014	022	010	012	082	068
15	103	094	099	094	072	067	070	066	025	044	023	022	082	068
14	107	098	102	095	081	089	091	064	050	042	037	034	082	071
13	103	115	072	086	108	096	089	076	078	075	049	058	090	054
12	110	111	080	070	106	098	085	085	111	089	074	066	064	070
11	107	085	052	084	132	111	085	078	135	110	100	085	060	067
10	079	080	043	066	099	116	083	074	139	132	125	105	049	050
9	068	065	042	047	085	092	072	070	133	123	119	121	045	056
8	041	152	017	038	077	067	070	070	104	108	114	127	032	041
7	023	036	022	025	043	061	065	062	081	099	116	124	022	027
6	014	019	008	017	021	038	038	064	060	066	093	089	013	025
5	005	012	006	009	012	020	042	053	032	039	060	077	008	015
4	002	010	002	006	010	009	028	036	014	015	042	034	006	013
3	001	002		002	002	002	017	031	005	007	023	025	001	002
2				*001		001	010	009	001	006	009	010		002
1					001	001	003	001			001	003	001	
N	1285	1169	1266	1185	1226	1167	1236	1161	1252	1192	1250	1162	1274	1206
M	13.0	12.6	14.6	13.6	11.8	11.5	11.2	10.8	10.0	9.9	8.9	8.7	14.8	14.5
S. D.	3.27	3.51	3.43	3.70	3.33	3.43	4.05	4.32	2.85	3.11	3.05	3.13	4.43	4.89

Appendix C. 3
Score Distributions of Criterion Quizzes (proportions)
Unit III

	M_1	M_2	N_1	N_2	O_1	O_2	P_1	P_2	Q_1	Q_2	R_1	R_2	U_1	U_2
24													025	002
23													044	006
22													040	023
21													035	019
20	002	007	*001										054	033
19	011	007	006		004	005							047	050
18	020	025	005	001	005	010			010				052	040
17	030	026	009	010	016	015			012	001		002	079	063
16	051	036	018	013	054	043	011	001	027	001	008	002	082	094
15	059	060	027	035	051	090	012	002	024	007	014	002	101	092
14	054	065	046	056	106	113	028	019	029	009	025	008	067	086
13	083	075	074	076	132	142	021	032	053	029	025	011	084	089
12	092	075	095	100	150	142	046	049	075	059	025	041	067	088
11	092	103	099	115	149	125	072	043	083	088	064	045	052	072
10	097	104	133	119	116	094	101	093	109	125	095	052	050	056
9	086	077	110	117	078	080	138	122	131	139	122	095	035	050
8	084	088	124	119	051	059	169	162	124	143	097	093	022	030
7	077	088	093	085	042	037	151	163	109	135	111	097	040	053
6	068	057	079	063	025	021	105	148	083	123	103	127	012	026
5	042	044	049	048	013	013	079	093	058	062	122	149	010	009
4	026	027	020	030	003	005	047	048	046	042	081	131	002	009
3	018	026	009	007	005	001	019	017	027	025	064	095		006
2	006	008	003	004		004	001	007		012	030	037		002
1	001	002		002		001		001			014	011		002
0												002		
N	1332	1129	1178	1072	765	1007	723	838	412	762	360	464	404	694
M	10.5	10.4	9.7	9.7	11.5	11.7	8.3	7.9	9.2	8.2	7.4	6.6	15.1	13.5
S.D.	3.83	3.88	3.11	3.05	2.78	3.79	2.75	2.60	3.21	2.61	3.19	2.94	4.61	4.26

Appendix D

Attitude Scale

Scale Value	Statement
8	Some parts of math are pretty bad, but on the whole I like it.
4	The things they teach in math courses are not practical enough.
1	Math isn't of any value and I don't intend to study it.
3	Studying math is all right, but only because if I don't I get in trouble with my parents and the teacher.
5	I don't expect to get a better grade than C in math.
9	I would like to know more about math than I do.
7	I could do better in math if I liked it better.
11	I would rather study math than any other subject.
10	Math homework is fun to do.
2	I don't understand how people can like math.
6	I have no particular like or dislike for mathematics.
11	I'd like to be a mathematician when I finish school.
2	I wish there were some other course I could take in place of math.
10	I look forward to math class.
4	There's no use in learning a lot of things about math that I'll never use.
6	I'm not enthusiastic about math, but I don't really dislike it either.
2	I hate math.
8	I know of only two or three subjects that I like better than math.
7	I sort of like mathematics.
5	My math grades are good enough but I don't like the subject very well.
9	The more you get to know about math, the better you like it.
3	I don't think I'll ever get interested in math.